

11/05/99

U.S. PTO

UTILITY PATENT APPLICATION TRANSMITTAL

Only for new nonprovisional applications under 37 CFR 1.53(b))

Attorney Docket No.

2318-247

Total Pages

First Named Inventor or Application Identifier

Sean V. TAVTIGIAN

Express Mail Label No.

U.S. PTO
09/434882
11/05/99

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

1. ☐ Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification Total pages [86]
(preferred arrangement set forth below)
 - Descriptive title of the invention
 - Cross references to Related Applications
 - Statement Regarding Fed sponsored R&D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings
 - Detailed Description
 - Claims
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 USC 113) (Total Sheets) [3]
4. ☐ Oath or Declaration (Total Pages) []
 - a. ☐ Newly executed (original or copy)
 - b. ☐ Copy from a prior application
(37 CFR 1.63(d)
(for continuation/divisional with Box 17 completed)

[Note Box 5 below]**☐ DELETION OF INVENTOR(S)**

Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b)

5. ☐ Incorporation by Reference (useable if Box 4b is checked) The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.

ADDRESS TO: **Assistant Commissioner of Patents**
Box Patent Application
Washington, D.C. 20231

6. ☐ Microfiche Computer Program (Appendix)
7. Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
 - a. ☒ Computer Readable Copy
 - b. ☒ Paper Copy (identical to computer copy)
Total pages [61]
 - c. ☒ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & documents)
9. ☐ 37 CFR 3.73(b) Statement
(when there is an assignee)
- ☐ Power of Attorney
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement /PTO 1449
- ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
14. ☐ Small Entity Statement(s)
- ☐ Statement Filed in prior application,
Status still proper and desired
15. ☐ Certified Copy of Priority Document(s).
(if foreign priority is claimed)
16. ☐ Other:

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.:

18. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label or ☒ Correspondence address below
 (Insert Customer No. or Attach bar code label here)

Name	Jeffrey L. Ihnen, Reg. No. 28,957 <i>Jeffrey Ihnen</i>				
Address	Rothwell, Figg, Ernst & Kurz Suite 701-East, 555 13th Street, N.W.				
City	Washington	State	D.C.	Zip Code	20004
Country	U.S.A.	Telephone	202-783-6040	Fax	202-783-6031

TITLE OF THE INVENTION

CHROMOSOME 17p-LINKED PROSTATE CANCER SUSCEPTIBILITY GENE

CROSS REFERENCE TO RELATED APPLICATION

- 5 The present invention is related to U.S. provisional patent application Serial No. 60/107,468, filed 6 November 1998, incorporated herein by reference.

BACKGROUND OF THE INVENTION

- 10 The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC2), some mutant alleles of which cause susceptibility to cancer, in particular, prostate cancer. More specifically, the invention relates to germline mutations in the HPC2 gene and their use in the diagnosis of predisposition to prostate cancer. The present invention further relates to somatic mutations in the HPC2 gene in human prostate cancer and their use in the diagnosis and prognosis of human prostate cancer. The invention also relates to germline mutations in the HPC2 gene and their use in the diagnosis of predisposition to other human cancers. Additionally, the invention relates to somatic mutations in the HPC2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC2 gene, including gene therapy, protein replacement therapy and protein mimetics. The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC2 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer.

- 25 The publications and other materials used herein to illuminate the background of the invention, and in particular, cases to provide additional details respecting the practice, are incorporated herein by reference, and for convenience, are referenced by author and date in the following text and respectively grouped in the appended List of References.

- 30 The genetics of cancer is complicated, involving the function of three loosely defined classes of genes: (1) dominant, positive regulators of the transformed state (oncogenes); (2) recessive, negative regulators of the transformed state (tumor suppressor genes); (3) genes that modify risk without playing a direct role in the biology of transformed cells (risk modifiers).

 Specific germline alleles of certain oncogenes and tumor suppressor genes are causally associated with predisposition to cancer. This set of genes is referred to as tumor predisposition genes. Some of the tumor predisposition genes which have been cloned and characterized

influence susceptibility to: 1) Retinoblastoma (RB1); 2) Wilms' tumor (WT1); 3) Li-Fraumeni (TP53); 4) Familial adenomatous polyposis (APC); 5) Neurofibromatosis type 1 (NF1); 6) Neurofibromatosis type 2 (NF2); 7) von Hippel-Lindau syndrome (VHL); 8) Multiple endocrine neoplasia type 2A (MEN2A); 9) Melanoma (CDKN2 and CDK4); 10) Breast and
 5 ovarian cancer (BRCA1 and BRCA2); 11) Cowden disease (MMAC1); 12) Multiple endocrine neoplasia (MEN1); 13) Nevoid basal cell carcinoma syndrome (PTC); 14) Tuberous sclerosis 2 (TSC2); 15) Xeroderma pigmentosum (genes involved in nucleotide excision repair); 16) Hereditary nonpolyposis colorectal cancer (genes involved in mismatch repair).

Specific germline alleles of certain risk modifier genes are also associated with
 10 predisposition to cancer, but the increased risk is sometimes only clearly expressed when it is combined with certain environmental, dietary, or other factors. Alcohol dehydrogenase (ADH) oxidizes ethanol to acetaldehyde, a chemical which is both mutagenic and carcinogenic in lab animals. The enzyme encoded by the ADH3¹ allele oxidizes ethanol relatively quickly, whereas the enzyme encoded by the ADH3² allele oxidizes ethanol more slowly. ADH3¹ homozygotes
 15 presumably have a high capacity for synthesis of acetaldehyde; those who also drink heavily are at increased risk for oral cavity, esophageal, and (in women) breast cancer relative to ADH3² homozygotes who drink equally heavily (Harty et al., 1997; Hori et al., 1997; Shields, 1997). The acetyltransferases encoded by N-acetyltransferase 1 (NAT1) and N-acetyltransferase 2 (NAT2) catalyze the acetylation of numerous xenobiotics including the aromatic amine
 20 carcinogens derived from smoking tobacco products. Individuals who are homozygous for slow acetylating forms of NAT2 who are also heavy smokers are at greater risk for lung, bladder, and (in females) breast cancer than individuals who smoke equally heavily but are homozygous for fast acetylating forms of NAT2 (Shields, 1997; Bouchardy et al., 1998).

The risk of hormone related cancers such as breast and prostate cancer may be modulated
 25 by allelic variants in enzymes that play a role in estrogen or androgen metabolism, or variants in proteins that mediate the biological effects of estrogens or androgens. A polymorphic CAG repeat in the human androgen receptor gene encodes a polymorphic polyglutamine tract near the amino-terminus of the protein. The length of the polyglutamine tract is inversely correlated with the transcriptional activation activity of the androgen receptor and thus one aspect of the
 30 biological response to androgens. Men whose androgen receptor contains a relatively short polyglutamine tract are at higher risk for prostate cancer, especially high stage/high histologic grade prostate cancer, than men whose androgen receptor contains a relatively long polyglutamine tract (Giovannucci et al., 1997).

Prostate cancer is the most common cancer in men in many western countries, and the second leading cause of cancer deaths in men. It accounts for more than 40,000 deaths in the US annually. The number of deaths is likely to continue rising over the next 10 to 15 years. In the US, prostate cancer is estimated to cost \$1.5 billion per year in direct medical expenses. In addition to the burden of suffering, it is a major public-health issue. Numerous studies have provided evidence for familial clustering of prostate cancer, indicating that family history is a major risk factor for this disease (Cannon et al., 1982; Steinberg et al., 1990; Carter et al., 1993).

Prostate cancer has long been recognized to be, in part, a familial disease. Numerous investigators have examined the evidence for genetic inheritance and concluded that the data are most consistent with dominant inheritance for a major susceptibility locus or loci. Woolf (1960), described a relative risk of 3.0 of developing prostate cancer among first-degree relatives of prostate cancer cases in Utah using death certificate data. Relative risks ranging from 3 to 11 for first-degree relatives of prostate cancer cases have been reported (Cannon et al., 1982; Woolf, 1960; Fincham et al., 1990; Meikle et al., 1985; Krain, 1974; Morganti et al., 1956; Goldgar et al., 1994). Carter et al. (1992) performed segregation analysis on families ascertained through a single prostate cancer proband. The analysis suggested Mendelian inheritance in a subset of families through autosomal dominant inheritance of a rare ($q=0.003$), high-risk allele with estimated cumulative risk of prostate cancer for carriers of 88% by age 85. Inherited prostate cancer susceptibility accounted for a significant proportion of early-onset disease, and overall was responsible for 9% of prostate occurrence by age 85. Recent results demonstrate that at least four loci exist which convey susceptibility to prostate cancer as well as other cancers. These loci are HPC1 on chromosome 1q, (Smith et al., 1996), HPC2 on chromosome 17p (this patent application, Example 1), HPCX on chromosome Xp (Xu et al., 1998), and one or more loci responsible for the unmapped residual.

Detection of genetic linkage for prostate cancer susceptibility to a defined segment of a chromosome requires that DNA sequence variants within that chromosomal segment confer the cancer susceptibility. This is usually taken to mean that the causal sequence variant(s) will either alter the expression of one or more linked genes or will alter the function of one of the linked genes. However, detection of the genetic linkage does not necessarily provide evidence for what class of gene (i.e. tumor suppressor, oncogene, or risk modifier) is affected by the causal sequence variant(s).

Most strategies for proceeding from genetic linkage of prostate cancer susceptibility to chromosome 17p to identification of the 17p-linked prostate cancer predisposing gene (HPC2)

require precise genetic localization studies to define a discrete segment of the chromosome within which the causal sequence variant(s) must map. Gene identification projects based on precise genetic localization are called positional cloning projects. The general strategy in positional cloning is to find all of the genes located within the genetically defined interval, identify sequence variants in and around those genes, and then determine which of those sequence variants either alter the expression or the function of one (or more) of the associated genes. Segregation of such sequence variants with the disease in the linked kindreds must also be demonstrated. We have executed a positional cloning project in the HPC2 region of chromosome 17p and found a gene, herein named HPC2, germline mutations which predispose individuals to prostate cancer.

SUMMARY OF THE INVENTION

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC2), some alleles of which cause susceptibility to cancer, in particular prostate cancer. More specifically, the present invention relates to germline mutations in the HPC2 gene and their use in the diagnosis of predisposition to prostate cancer. The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPC2 gene. The invention further relates to somatic mutations in the HPC2 gene in human prostate cancer and their use in the diagnosis and prognosis of human prostate cancer. Additionally, the invention relates to somatic mutations in the HPC2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC2 gene, (including gene therapy, protein replacement therapy, protein mimetics, and inhibitors). The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC2 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer. The HPC2 gene is useful as a marker for the HPC2 locus and as a marker for prostate cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a multipoint linkage analysis of 4 kindreds that show suggestive evidence for linkage to the HPC2 prostate cancer susceptibility locus relative to chromosome 17p markers.

Figures 2A-B are diagrams showing the order of genetic markers and recombinant boundaries neighboring HPC2, a schematic map of BACs spanning the HPC2 region, a schematic map of transcription units within the HPC2 region, and two diagrams of the HPC2 transcription unit showing the locations of the exons of HPC2 relative to the BAC to which it maps and relative to each other. The individual exons are numbered.

Figure 3 is an alignment of the sequence of exon 1 of the human HPC2 gene with exon 1 of the mouse HPC2 gene. The figure also shows an alignment of the peptide sequence encoded by exon 1 of the human HPC2 gene with the peptide sequence encoded by exon 1 of the mouse HPC2 gene. The human DNA sequence is SEQ ID NO:210; the human amino acid sequence is SEQ ID NO:211; the mouse DNA sequence is SEQ ID NO:212 and the mouse amino acid sequence is SEQ ID NO:213.

BRIEF DESCRIPTION OF THE TABLES

Table 1 is a compilation of 2-point LOD scores for markers in the HPC2 region.

Table 2 is a list of the accession numbers of human EST sequences used to assemble a tentative, partial cDNA sequence of the human HPC2 gene.

Table 3 is a list of the primers used for obtaining 5' RACE products that contained the start codon and part of the 5' UTR of the human HPC2 gene, primers used to prepare a full length human HPC2 expression construct, and primers used to check the sequence of that construct.

Table 4 is a list of the accession numbers of mouse EST sequences used to assemble a tentative, partial cDNA sequence of the mouse HPC2 gene.

Table 5 is a list of the primers used for obtaining 5' RACE products that contained the start codon and part of the 5' UTR of the mouse HPC2 gene, primers used to prepare a full length mouse HPC2 expression construct, and primers used to check the sequence of that construct.

Table 6 is a list of the primers used to mutation screen the human HPC2 gene from genomic DNA.

Table 7 is a summary of germline sequence variants of the human HPC2 gene.

SUMMARY OF SEQUENCE LISTING

SEQ ID NO:1 is the nucleotide sequence for the human HPC2 cDNA from the start codon through the stop codon. SEQ ID NO:2 is the amino acid sequence for the human HPC2

protein. SEQ ID NO:3 is the nucleotide sequence for the human HPC2 cDNA from 50 base pairs before the start codon through the end of the 3' UTR. SEQ ID NO:4 to SEQ ID NO:27 are the sequences of exon 1 to exon 24 of the human HPC2 gene. SEQ ID NO:28 is the genomic sequence of the human HPC2 gene. SEQ ID NOs:29-190 are nucleotide sequences of primers used to identify the human and/or mouse HPC2 genes or to screen for mutations. SEQ ID NOs:191-209 are nucleotide sequences of the HPC2 around and including various sequence variants. SEQ ID NO:210 is the nucleotide sequence of human HPC2 exon 1 and SEQ ID NO:211 is the corresponding amino acid sequence. SEQ ID NO:213 is nucleotide sequence of mouse HPC2 exon 1 and SEQ ID NO:213 is the corresponding amino acid sequence.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides an isolated polynucleotide comprising all, or a portion of the HPC2 locus or of a mutated HPC2 locus, preferably at least eight bases and not more than about 27 kb in length. Such polynucleotides may be antisense polynucleotides. The present invention also provides a recombinant construct comprising such an isolated polynucleotide, for example, a recombinant construct suitable for expression in a transformed host cell.

Also provided by the present invention are methods of detecting a polynucleotide comprising a portion of the HPC2 locus or its expression product in an analyte. Such methods may further comprise the step of amplifying the portion of the HPC2 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the HPC2 locus. The method is useful for either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer. The HPC2 gene is useful as a marker for the HPC2 locus and as a marker for prostate cancer.

The present invention also provides isolated antibodies, preferably monoclonal antibodies, which specifically bind to an isolated polypeptide comprised of at least five amino acid residues encoded by the HPC2 locus.

The present invention also provides kits for detecting in an analyte a polynucleotide comprising a portion of the HPC2 locus, the kits comprising a polynucleotide complementary to the portion of the HPC2 locus packaged in a suitable container, and instructions for its use.

The present invention further provides methods of preparing a polynucleotide comprising polymerizing nucleotides to yield a sequence comprised of at least eight consecutive nucleotides of the HPC2 locus; and methods of preparing a polypeptide comprising poly-

merizing amino acids to yield a sequence comprising at least five amino acids encoded within the HPC2 locus.

The present invention further provides methods of screening the HPC2 gene to identify mutations. Such methods may further comprise the step of amplifying a portion of the HPC2 locus, and may further include a step of providing a set of polynucleotides which are primers for amplification of said portion of the HPC2 locus. Such methods may also include a step of providing the complete set of short polynucleotides defined by the sequence of HPC2 or discrete subsets of that sequence, all single-base substitutions of that sequence or discrete subsets of that sequence, all 1-, 2-, 3-, or 4-base deletions of that sequence or discrete subsets of that sequence, and all 1-, 2-, 3-, or 4-base insertions in that sequence or discrete subsets of that sequence. The method is useful for identifying mutations for use in either diagnosis of the predisposition to cancer or the diagnosis or prognosis of cancer.

The present invention further provides methods of screening suspected HPC2 mutant alleles to identify mutations in the HPC2 gene.

In addition, the present invention provides methods to screen drugs for inhibition or restoration of HPC2 gene product function as an anticancer therapy.

Finally, the present invention provides the means necessary for production of gene-based therapies directed at cancer cells. These therapeutic agents may take the form of polynucleotides comprising all or a portion of the HPC2 locus placed in appropriate vectors or delivered to target cells in more direct ways such that the function of the HPC2 protein is reconstituted. Therapeutic agents may also take the form of polypeptides based on either a portion of, or the entire protein sequence of HPC2. These may functionally replace the activity of HPC2 in vivo.

It is a discovery of the present invention that the HPC2 locus which predisposes individuals to prostate cancer, is a gene encoding an HPC2 protein, which has been found to be non-identical to publicly available protein or cDNA sequences. This gene is termed HPC2 herein. It is a discovery of the present invention that mutations in the HPC2 locus in the germline are indicative of a predisposition to prostate cancer. Finally, it is a discovery of the present invention that germline mutations in the HPC2 locus are also associated with prostate cancer and other types of cancer. The mutational events of the HPC2 locus can involve deletions, insertions and nucleotide substitutions within the coding sequence and the non-coding sequence.

Useful Diagnostic Techniques

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type HPC2 locus is detected. In addition, the method can be performed by detecting the wild-type HPC2 locus and confirming the lack of a predisposition to cancer at the HPC2 locus. "Alteration of a wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or of only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, and are not inherited in the germline. Germline mutations can be found in any of a body's tissues and are inherited. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are somatically mutated, then a late neoplastic state is indicated. The finding of HPC2 mutations thus provides both diagnostic and prognostic information. An HPC2 allele which is not deleted (e.g., found on the sister chromosome to a chromosome carrying an HPC2 deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many mutations found in tumor tissues will be those leading to decreased expression of the HPC2 gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to reduction or loss of expression of the HPC2 gene product, expression of an altered HPC2 gene product, or to a decrease in mRNA stability or translation efficiency.

Useful diagnostic techniques include, but are not limited to fluorescent *in situ* hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNase protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, hybridization using nucleic acid modified with gold nanoparticles and PCR-SSCP, as discussed in detail further below. Also useful is the recently developed technique of DNA microchip technology.

Predisposition to cancers, such as prostate cancer, and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the HPC2 gene. For example, a person who has inherited a germline HPC2 mutation would be prone to develop cancers. This can be determined by testing DNA from any tissue of the person's body. Most

simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic cells for mutations of the HPC2 gene. Alteration of a wild-type HPC2 allele, whether, for example, by point mutation or deletion, can be detected by any of the means discussed herein.

5 There are several methods that can be used to detect DNA sequence variation. Direct DNA sequencing, either manual sequencing or automated fluorescent sequencing can detect sequence variation. For a gene as large as HPC2, manual sequencing is very labor-intensive, but under optimal conditions, mutations in the coding sequence of a gene are rarely missed. Another approach is the single-stranded conformation polymorphism assay (SSCA) (Orita *et al.*, 10 1989). This method does not detect all sequence changes, especially if the DNA fragment size is greater than 200 bp, but can be optimized to detect most DNA sequence variation. The reduced detection sensitivity is a disadvantage, but the increased throughput possible with SSCA makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. The fragments which have shifted mobility on SSCA gels are then sequenced to 15 determine the exact nature of the DNA sequence variation. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE) (Sheffield *et al.*, 1991), heteroduplex analysis (HA) (White *et al.*, 1992) and chemical mismatch cleavage (CMC) (Grompe *et al.*, 1989). None of the methods described above will detect large deletions, duplications or insertions, nor will they 20 detect a regulatory mutation which affects transcription or translation of the protein. Other methods which might detect these classes of mutations such as a protein truncation assay or the asymmetric assay, detect only specific types of mutations and would not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a recent review by Grompe (1993). Once a mutation is known, an allele specific 25 detection approach such as allele specific oligonucleotide (ASO) hybridization can be utilized to rapidly screen large numbers of other samples for that same mutation. Such a technique can utilize probes which are labeled with gold nanoparticles to yield a visual color result (Elghanian *et al.*, 1997).

30 In order to detect the alteration of the wild-type HPC2 gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These techniques, as well as other techniques for separating tumor cells from normal cells, are

well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

Detection of point mutations may be accomplished by molecular cloning of the HPC2 allele(s) and sequencing the allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified directly from a genomic DNA preparation from the tumor tissue, using known techniques. The DNA sequence of the amplified sequences can then be determined.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single-stranded conformation analysis (SSCA) (Orita *et al.*, 1989); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell *et al.*, 1990; Sheffield *et al.*, 1989); 3) RNase protection assays (Finkelstein *et al.*, 1990; Kinszler *et al.*, 1991); 4) allele-specific oligonucleotides (ASOs) (Conner *et al.*, 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Ruano and Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a particular HPC2 mutation. If the particular HPC2 mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton *et al.*, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the HPC2 mutation found in that individual. Other techniques for detecting insertions and deletions as known in the art can be used.

In the first three methods (SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein binds

only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or in its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An example of a mismatch cleavage technique is the RNase protection method. In the practice of the present invention, the method involves the use of a labeled riboprobe which is complementary to the human wild-type HPC2 gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the HPC2 mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the HPC2 mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton *et al.*, 1988; Shenk *et al.*, 1975; Novack *et al.*, 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the HPC2 gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the HPC2 gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the HPC2 gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length (although shorter and longer oligomers are also usable as well recognized by those of skill in the art), corresponding to a portion of the HPC2 gene sequence. By use of a battery of such allele-specific probes, PCR amplification products

can be screened to identify the presence of a previously identified mutation in the HPC2 gene. Hybridization of allele-specific probes with amplified HPC2 sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under high stringency hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

The newly developed technique of nucleic acid analysis via microchip technology is also applicable to the present invention. In this technique, literally thousands of distinct oligonucleotide probes are built up in an array on a silicon chip. Nucleic acid to be analyzed is fluorescently labeled and hybridized to the probes on the chip. It is also possible to study nucleic acid-protein interactions using these nucleic acid microchips. Using this technique one can determine the presence of mutations or even sequence the nucleic acid being analyzed or one can measure expression levels of a gene of interest. The method is one of parallel processing of many, even thousands, of probes at once and can tremendously increase the rate of analysis. Several papers have been published which use this technique. Some of these are Hacia et al., 1996; Shoemaker et al., 1996; Chee et al., 1996; Lockhart et al., 1996; DeRisi et al., 1996; Lipshutz et al., 1995. This method has already been used to screen people for mutations in the breast cancer gene BRCA1 (Hacia et al., 1996). This new technology has been reviewed in a news article in Chemical and Engineering News (Borman, 1996) and been the subject of an editorial (Nature Genetics, 1996). Also see Fodor (1997).

The most definitive test for mutations in a candidate locus is to directly compare genomic HPC2 sequences from cancer patients with those from a control population. Alternatively, one could sequence messenger RNA after amplification, e.g., by PCR, thereby eliminating the necessity of determining the exon structure of the candidate gene.

Mutations from cancer patients falling outside the coding region of HPC2 can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the HPC2 gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of HPC2 mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type HPC2 gene. Alteration of wild-type HPC2 genes can also be detected by screening for alteration of wild-type HPC2 protein. For example, monoclonal antibodies immunoreactive with HPC2 can be used to screen a tissue.

Lack of cognate antigen would indicate an HPC2 mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant HPC2 gene product. Such immunological assays can be done in any convenient formats known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered HPC2 protein can be used to detect alteration of wild-type HPC2 genes. Functional assays, such as protein binding determinations, can be used. In addition, assays can be used which detect HPC2 biochemical function. Finding a mutant HPC2 gene product indicates alteration of a wild-type HPC2 gene.

Mutant HPC2 genes or gene products can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant HPC2 genes or gene products in tissues can be applied to other body samples. Cancer cells are sloughed off from tumors and appear in such body samples. In addition, the HPC2 gene product itself may be secreted into the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant HPC2 genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which HPC2 has a role in tumorigenesis. The diagnostic method of the present invention is useful for clinicians, so they can decide upon an appropriate course of treatment.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular HPC2 allele using PCR. The pairs of single-stranded DNA primers can be annealed to sequences within or surrounding the HPC2 gene on chromosome 17 in order to prime amplifying DNA synthesis of the HPC2 gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the HPC2 gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele-specific primers can also be used. Such primers anneal only to particular HPC2 mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from HPC2 sequences or sequences adjacent to HPC2, except for the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in

the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the HPC2 open reading frame shown in SEQ ID NOs:1 and 3, design of particular primers is well within the skill of the art.

5 The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR amplification products. They may also be used to detect mismatches with the HPC2 gene or mRNA using other techniques.

10 It has been discovered that individuals with the wild-type HPC2 gene do not have cancer which results from the HPC2 allele. However, mutations which interfere with the function of the HPC2 protein are involved in the pathogenesis of cancer. Thus, the presence of an altered (or a mutant) HPC2 gene which produces a protein having a loss of function, or altered function, directly correlates to an increased risk of cancer. In order to detect an HPC2 gene mutation, a biological sample is prepared and analyzed for a difference between the sequence of the HPC2
15 allele being analyzed and the sequence of the wild-type HPC2 allele. Mutant HPC2 alleles can be initially identified by any of the techniques described above. The mutant alleles are then sequenced to identify the specific mutation of the particular mutant allele. Alternatively, mutant HPC2 alleles can be initially identified by identifying mutant (altered) HPC2 proteins, using conventional techniques. The mutant alleles are then sequenced to identify the specific mutation
20 for each allele. The mutations, especially those which lead to an altered function of the HPC2 protein, are then used for the diagnostic and prognostic methods of the present invention.

Definitions

The present invention employs the following definitions:

"**Amplification of Polynucleotides**" utilizes methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. Also useful are strand displacement amplification (SDA), thermophilic SDA, and nucleic acid sequence based amplification (3SR or NASBA). These methods are well known and widely practiced in the art. See, e.g., U.S. Patents 4,683,195 and 4,683,202 and Innis *et al.*, 1990 (for PCR); and Wu and Wallace, 1989 (for LCR); U.S. Patents 5,270,184 and 5,455,166 and Walker *et al.*, 1992 (for SDA); Spargo *et al.*, 1996 (for thermophilic SDA) and U.S. Patent 5,409,818, Fahy *et al.*, 1991 and Compton, 1991 for 3SR and NASBA. Reagents and hardware for conducting PCR are commercially available. Primers useful to amplify sequences from the HPC2 region are preferably complementary to, and hybridize specifically to sequences in the HPC2 region or in regions that flank a target region therein. HPC2 sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

"**Analyte polynucleotide**" and "**analyte strand**" refer to a single- or double-stranded polynucleotide which is suspected of containing a target sequence, and which may be present in a variety of types of samples, including biological samples.

"**Antibodies.**" The present invention also provides polyclonal and/or monoclonal antibodies and fragments thereof, and immunologic binding equivalents thereof, which are capable of specifically binding to the HPC2 polypeptides and fragments thereof or to polynucleotide sequences from the HPC2 region, particularly from the HPC2 locus or a portion thereof. The term "**antibody**" is used both to refer to a homogeneous molecular entity, or a mixture such as a serum product made up of a plurality of different molecular entities. Polypeptides may be prepared synthetically in a peptide synthesizer and coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected over several months into rabbits. Rabbit sera is tested for immunoreactivity to the HPC2 polypeptide or fragment. Monoclonal antibodies may be made by injecting mice with the protein polypeptides, fusion proteins or fragments thereof. Monoclonal antibodies will be screened by ELISA and tested for specific immunoreactivity with HPC2 polypeptide or fragments thereof. See, Harlow and Lane, 1988. These antibodies will be useful in assays as well as pharmaceuticals.

Once a sufficient quantity of desired polypeptide has been obtained, it may be used for various purposes. A typical use is the production of antibodies specific for binding. These

antibodies may be either polyclonal or monoclonal, and may be produced by *in vitro* or *in vivo* techniques well known in the art. For production of polyclonal antibodies, an appropriate target immune system, typically mouse or rabbit, is selected. Substantially purified antigen is presented to the immune system in a fashion determined by methods appropriate for the animal and by other parameters well known to immunologists. Typical sites for injection are in footpads, intramuscularly, intraperitoneally, or intradermally. Of course, other species may be substituted for mouse or rabbit. Polyclonal antibodies are then purified using techniques known in the art, adjusted for the desired specificity.

An immunological response is usually assayed with an immunoassay. Normally, such immunoassays involve some purification of a source of antigen, for example, that produced by the same cells and in the same fashion as the antigen. A variety of immunoassay methods are well known in the art. See, e.g., Harlow and Lane, 1988, or Goding, 1986.

Monoclonal antibodies with affinities of 10^{-8} M⁻¹ or preferably 10^{-9} to 10^{-10} M⁻¹ or stronger will typically be made by standard procedures as described, e.g., in Harlow and Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse *et al.*, 1989. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, polypeptides and antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced (see U.S. Patent 4,816,567).

"Binding partner" refers to a molecule capable of binding a ligand molecule with high specificity, as for example, an antigen and an antigen-specific antibody or an enzyme and its

inhibitor. In general, the specific binding partners must bind with sufficient affinity to immobilize the analyte copy/complementary strand duplex (in the case of polynucleotide hybridization) under the isolation conditions. Specific binding partners are known in the art and include, for example, biotin and avidin or streptavidin, IgG and protein A, the numerous, known
 5 receptor-ligand couples, and complementary polynucleotide strands. In the case of complementary polynucleotide binding partners, the partners are normally at least about 15 bases in length, and may be at least 40 bases in length. It is well recognized by those of skill in the art that lengths shorter than 15 (e.g., 8 bases), between 15 and 40, and greater than 40 bases may also be used. The polynucleotides may be composed of DNA, RNA, or synthetic
 10 nucleotide analogs. Further binding partners can be identified using, e.g., the two-hybrid yeast screening assay as described herein.

A **"biological sample"** refers to a sample of tissue or fluid suspected of containing an analyte polynucleotide or polypeptide from an individual including, but not limited to, e.g., plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal,
 15 and genitourinary tracts, tears, saliva, blood cells, tumors, organs, tissue and samples of *in vitro* cell culture constituents.

As used herein, the terms **"diagnosing"** or **"prognosing,"** as used in the context of neoplasia, are used to indicate 1) the classification of lesions as neoplasia, 2) the determination of the severity of the neoplasia, or 3) the monitoring of the disease progression, prior to, during
 20 and after treatment.

"Encode". A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be
 25 deduced therefrom.

"Isolated" or **"substantially pure"**. An "isolated" or "substantially pure" nucleic acid (e.g., an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components which naturally accompany a native human sequence or protein, e.g., ribosomes, polymerases, many other human genome sequences and proteins. The term
 30 embraces a nucleic acid sequence or protein which has been removed from its naturally occurring environment, and includes recombinant or cloned DNA isolates and chemically synthesized analogs or analogs biologically synthesized by heterologous systems.

"HPC2 Allele" refers to normal alleles of the HPC2 locus as well as alleles carrying variations that predispose individuals to develop prostate cancer. Such predisposing alleles are also called "HPC2 susceptibility alleles".

"HPC2 Locus", "HPC2 Gene", "HPC2 Nucleic Acids" or "HPC2 Polynucleotide" each refer to polynucleotides, all of which are in the HPC2 region, that are likely to be expressed in normal tissue, certain alleles of which predispose an individual to develop prostate cancers. Mutations at the HPC2 locus may be involved in the initiation and/or progression of other types of tumors. The locus is indicated in part by mutations that predispose individuals to develop cancer. These mutations fall within the HPC2 region described *infra*. The HPC2 locus is intended to include coding sequences, intervening sequences and regulatory elements controlling transcription and/or translation. The HPC2 locus is intended to include all allelic variations of the DNA sequence.

These terms, when applied to a nucleic acid, refer to a nucleic acid which encodes an HPC2 polypeptide, fragment, homolog or variant, including, e.g., protein fusions or deletions. The nucleic acids of the present invention will possess a sequence which is either derived from, or substantially similar to a natural HPC2-encoding gene or one having substantial homology with a natural HPC2-encoding gene or a portion thereof.

The HPC2 gene or nucleic acid includes normal alleles of the HPC2 gene, including silent alleles having no effect on the amino acid sequence of the HPC2 polypeptide as well as alleles leading to amino acid sequence variants of the HPC2 polypeptide that do not substantially affect its function. These terms also include alleles having one or more mutations which adversely affect the function of the HPC2 polypeptide. A mutation may be a change in the HPC2 nucleic acid sequence which produces a deleterious change in the amino acid sequence of the HPC2 polypeptide, resulting in partial or complete loss of HPC2 function, or may be a change in the nucleic acid sequence which results in the loss of effective HPC2 expression or the production of aberrant forms of the HPC2 polypeptide.

The HPC2 nucleic acid may be that shown in SEQ ID NOs:1, 3 or 28 or it may be an allele as described above or a variant or derivative differing from that shown by a change which is one or more of addition, insertion, deletion and substitution of one or more nucleotides of the sequence shown. Changes to the nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the genetic code.

Thus, nucleic acid according to the present invention may include a sequence different from the sequence shown in SEQ ID NOs:1, 3 or 28 yet encode a polypeptide with the same

amino acid sequence as shown in SEQ ID NO:1. That is, nucleic acids of the present invention include sequences which are degenerate as a result of the genetic code. On the other hand, the encoded polypeptide may comprise an amino acid sequence which differs by one or more amino acid residues from the amino acid sequence shown in SEQ ID NO:2. Nucleic acid
 5 encoding a polypeptide which is an amino acid sequence variant, derivative or allele of the amino acid sequence shown in SEQ ID NO:2 is also provided by the present invention.

The HPC2 gene also refers to (a) any DNA sequence that (i) hybridizes to the complement of the DNA sequences that encode the amino acid sequence set forth in SEQ ID NO:2 under highly stringent conditions (Ausubel et al., 1992) and (ii) encodes a gene product
 10 functionally equivalent to HPC2, or (b) any DNA sequence that (i) hybridizes to the complement of the DNA sequences that encode the amino acid sequence set forth in SEQ ID NO:2 under less stringent conditions, such as moderately stringent conditions (Ausubel et al., 1992) and (ii) encodes a gene product functionally equivalent to HPC2. The invention also includes nucleic acid molecules that are the complements of the sequences described herein.

The polynucleotide compositions of this invention include RNA, cDNA, genomic DNA, synthetic forms, and mixed polymers, both sense and antisense strands, and may be chemically or biochemically modified or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those skilled in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog,
 20 internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, etc.), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids, etc.). Also included are synthetic molecules that mimic polynucleotides in their ability to bind
 25 to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The present invention provides recombinant nucleic acids comprising all or part of the HPC2 region. The recombinant construct may be capable of replicating autonomously in a host
 30 cell. Alternatively, the recombinant construct may become integrated into the chromosomal DNA of the host cell. Such a recombinant polynucleotide comprises a polynucleotide of genomic, cDNA, semi-synthetic, or synthetic origin which, by virtue of its origin or manipulation, 1) is not associated with all or a portion of a polynucleotide with which it is

associated in nature; 2) is linked to a polynucleotide other than that to which it is linked in nature; or 3) does not occur in nature. Where nucleic acid according to the invention includes RNA, reference to the sequence shown should be construed as reference to the RNA equivalent, with U substituted for T.

5 Therefore, recombinant nucleic acids comprising sequences otherwise not naturally occurring are provided by this invention. Although the wild-type sequence may be employed, it will often be altered, e.g., by deletion, substitution or insertion.

cDNA or genomic libraries of various types may be screened as natural sources of the nucleic acids of the present invention, or such nucleic acids may be provided by amplification of
10 sequences resident in genomic DNA or other natural sources, e.g., by PCR. The choice of cDNA libraries normally corresponds to a tissue source which is abundant in mRNA for the desired proteins. Phage libraries are normally preferred, but other types of libraries may be used. Clones of a library are spread onto plates, transferred to a substrate for screening, denatured and probed for the presence of desired sequences.

15 The DNA sequences used in this invention will usually comprise at least about five codons (15 nucleotides), more usually at least about 7-15 codons, and most preferably, at least about 35 codons. One or more introns may also be present. This number of nucleotides is usually about the minimal length required for a successful probe that would hybridize specifically with an HPC2-encoding sequence. In this context, oligomers of as low as 8
20 nucleotides, more generally 8-17 nucleotides, can be used for probes, especially in connection with chip technology.

Techniques for nucleic acid manipulation are described generally, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Reagents useful in applying such techniques, such as restriction enzymes and the like, are widely known in the art and commercially available
25 from such vendors as New England BioLabs, Boehringer Mannheim, Amersham, Promega Biotech, U. S. Biochemicals, New England Nuclear, and a number of other sources. The recombinant nucleic acid sequences used to produce fusion proteins of the present invention may be derived from natural or synthetic sequences. Many natural gene sequences are obtainable from various cDNA or from genomic libraries using appropriate probes. See, GenBank,
30 National Institutes of Health.

"HPC2 Region" refers to a portion of human chromosome 17 bounded by the markers D17S947 and D17S799. This region contains the HPC2 locus, including the HPC2 gene.

As used herein, the terms "**HPC2 locus**", "**HPC2 allele**" and "**HPC2 region**" all refer to the double-stranded DNA comprising the locus, allele, or region, as well as either of the single-stranded DNAs comprising the locus, allele or region.

As used herein, a "**portion**" of the *HPC2* locus or region or allele is defined as having a minimal size of at least about eight nucleotides, or preferably about 15 nucleotides, or more preferably at least about 25 nucleotides, and may have a minimal size of at least about 40 nucleotides. This definition includes all sizes in the range of 8-40 nucleotides as well as greater than 40 nucleotides. Thus, this definition includes nucleic acids of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400, 500 nucleotides, or nucleic acids having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72, 121, etc., nucleotides), or nucleic acids having more than 500 nucleotides. The present invention includes all novel nucleic acids having at least 8 nucleotides derived from any of SEQ ID NOs:1 or 3-28, its complement or functionally equivalent nucleic acid sequences. The present invention does not include nucleic acids which exist in the prior art. That is, the present invention includes all nucleic acids having at least 8 nucleotides derived from any of SEQ ID NOs:1 or 3-28 with the proviso that it does not include nucleic acids existing in the prior art.

"**HPC2 protein**" or "**HPC2 polypeptide**" refers to a protein or polypeptide encoded by the *HPC2* locus, variants or fragments thereof. The term "polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Ordinarily, such polypeptides will be at least about 50% homologous to the native *HPC2* sequence, preferably in excess of about 90%, and more preferably at least about 95% homologous. Also included are proteins encoded by DNA which hybridize under high or low stringency conditions, to *HPC2*-encoding nucleic acids and closely related polypeptides or proteins retrieved by antisera to the *HPC2* protein(s).

An *HPC2* polypeptide may be that derived from any of the exons described herein which may be in isolated and/or purified form, free or substantially free of material with which it is naturally associated. The polypeptide may, if produced by expression in a prokaryotic cell or

produced synthetically, lack native post-translational processing, such as glycosylation. Alternatively, the present invention is also directed to polypeptides which are sequence variants, alleles or derivatives of an HPC2 polypeptide. Such polypeptides may have an amino acid sequence which differs from that derived from any of the exons described herein by one or more
 5 of addition, substitution, deletion or insertion of one or more amino acids. Preferred such polypeptides have HPC2 function.

Substitutional variants typically contain the exchange of one amino acid for another at one or more sites within the protein, and may be designed to modulate one or more properties of the polypeptide, such as stability against proteolytic cleavage, without the loss of other functions
 10 or properties. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. Preferred substitutions are ones which are conservative, that is, one amino acid is replaced with one of similar shape and charge. Conservative substitutions are well known in the art and typically include substitutions within the following groups: glycine, alanine; valine,
 15 isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and tyrosine, phenylalanine.

Certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules or binding sites on
 20 proteins interacting with an HPC2 polypeptide. Since it is the interactive capacity and nature of a protein which defines that protein's biological functional activity, certain amino acid substitutions can be made in a protein sequence, and its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. In making such changes, the hydrophobic index of amino acids may be considered. The importance of the hydrophobic amino acid index
 25 in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, 1982). Alternatively, the substitution of like amino acids can be made effectively on the basis of hydrophilicity. The importance of hydrophilicity in conferring interactive biological function of a protein is generally understood in the art (U.S. Patent 4,554,101). The use of the hydrophobic index or hydrophilicity in designing polypeptides is further discussed in
 30 U.S. Patent 5,691,198.

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acids, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues.

"Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

The term **peptide mimetic** or **mimetic** is intended to refer to a substance which has the essential biological activity of an HPC2 polypeptide. A peptide mimetic may be a peptide-containing molecule that mimics elements of protein secondary structure (Johnson et al., 1993). The underlying rationale behind the use of peptide mimetics is that the peptide backbone of proteins exists chiefly to orient amino acid side chains in such a way as to facilitate molecular interactions, such as those of antibody and antigen, enzyme and substrate or scaffolding proteins. A peptide mimetic is designed to permit molecular interactions similar to the natural molecule. A mimetic may not be a peptide at all, but it will retain the essential biological activity of a natural HPC2 polypeptide.

"Probes". Polynucleotide polymorphisms associated with HPC2 alleles which predispose to certain cancers or are associated with most cancers are detected by hybridization with a polynucleotide probe which forms a stable hybrid with that of the target sequence, under highly stringent to moderately stringent hybridization and wash conditions. If it is expected that the probes will be perfectly complementary to the target sequence, high stringency conditions will be used. Hybridization stringency may be lessened if some mismatching is expected, for example, if variants are expected with the result that the probe will not be completely complementary. Conditions are chosen which rule out nonspecific/adventitious bindings, that is, which minimize noise. (It should be noted that throughout this disclosure, if it is simply stated that "stringent" conditions are used that is meant to be read as "high stringency" conditions are used.) Since such indications identify neutral DNA polymorphisms as well as mutations, these indications need further analysis to demonstrate detection of an HPC2 susceptibility allele. An example of high stringency conditions is to hybridize to filter bound DNA in 0.5 M NaHPO₄, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C and to wash in 0.1xSSC/0.1% SDS at 68°C (Ausubel et al., 1992). Less stringent conditions, such as moderately stringent conditions, are defined as above but with the wash step being in 0.2xSSC/0.1% SDS at 42°C.

Probes for HPC2 alleles may be derived from the sequences of the HPC2 region, its cDNA, functionally equivalent sequences, or the complements thereof. The probes may be of any suitable length, which span all or a portion of the HPC2 region, and which allow specific hybridization to the HPC2 region. If the target sequence contains a sequence identical to that of the probe, the probes may be short, e.g., in the range of about 8-30 base pairs, since the hybrid

will be relatively stable under even highly stringent conditions. If some degree of mismatch is expected with the probe, i.e., if it is suspected that the probe will hybridize to a variant region, a longer probe may be employed which hybridizes to the target sequence with the requisite specificity.

5 The probes will include an isolated polynucleotide attached to a label or reporter molecule and may be used to isolate other polynucleotide sequences, having sequence similarity by standard methods. For techniques for preparing and labeling probes see, e.g., Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992. Other similar polynucleotides may be selected by using homologous polynucleotides. Alternatively, polynucleotides encoding these or similar
10 polypeptides may be synthesized or selected by use of the redundancy in the genetic code. Various codon substitutions may be introduced, e.g., by silent changes (thereby producing various restriction sites) or to optimize expression for a particular system. Mutations may be introduced to modify the properties of the polypeptide, perhaps to change ligand-binding affinities, interchain affinities, or the polypeptide degradation or turnover rate.

15 Probes comprising synthetic oligonucleotides or other polynucleotides of the present invention may be derived from naturally occurring or recombinant single- or double-stranded polynucleotides, or be chemically synthesized. Probes may also be labeled by nick translation, Klenow fill-in reaction, or other methods known in the art.

Portions of the polynucleotide sequence having at least about eight nucleotides, usually
20 at least about 15 nucleotides, and fewer than about 9 kb, usually fewer than about 1.0 kb, from a polynucleotide sequence encoding HPC2 are preferred as probes. This definition therefore includes probes of sizes 8 nucleotides through 9000 nucleotides. Thus, this definition includes probes of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400 or 500 nucleotides or probes having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72,
25 121, etc., nucleotides), or probes having more than 500 nucleotides. The probes may also be used to determine whether mRNA encoding HPC2 is present in a cell or tissue. The present invention includes all novel probes having at least 8 nucleotides derived from any of SEQ ID NOs:1 or 3-28 its complement or functionally equivalent nucleic acid sequences. The present invention does not include probes which exist in the prior art. That is, the present invention
30 includes all probes having at least 8 nucleotides derived from any of SEQ ID NOs:1 or 3-28 with the proviso that they do not include probes existing in the prior art.

Similar considerations and nucleotide lengths are also applicable to primers which may be used for the amplification of all or part of the *HPC2* gene. Thus, a definition for primers

includes primers of 8, 12, 15, 20, 25, 40, 60, 80, 100, 200, 300, 400, 500 nucleotides, or primers having any number of nucleotides within these ranges of values (e.g., 9, 10, 11, 16, 23, 30, 38, 50, 72, 121, etc. nucleotides), or primers having more than 500 nucleotides, or any number of nucleotides between 500 and 9000. The primers may also be used to determine whether mRNA
5 encoding *HPC2* is present in a cell or tissue. The present invention includes all novel primers having at least 8 nucleotides derived from the *HPC2* locus for amplifying the *HPC2* gene, its complement or functionally equivalent nucleic acid sequences. The present invention does not include primers which exist in the prior art. That is, the present invention includes all primers having at least 8 nucleotides with the proviso that it does not include primers existing in the
10 prior art.

"Protein modifications or fragments" are provided by the present invention for *HPC2* polypeptides or fragments thereof which are substantially homologous to primary structural sequence but which include, e.g., *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate unusual amino acids. Such modifications include, for example, acetylation,
15 carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, e.g., with radionuclides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P , ligands which bind to labeled antigens (e.g., antibodies), fluorophores, chemiluminescent agents, enzymes,
20 and antigens which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods of labeling polypeptides are well known in the art. See Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

Besides substantially full-length polypeptides, the present invention provides for
25 biologically active fragments of the polypeptides. Significant biological activities include ligand-binding, immunological activity and other biological activities characteristic of *HPC2* polypeptides. Immunological activities include both immunogenic function in a target immune system, as well as sharing of immunological epitopes for binding, serving as either a competitor or substitute antigen for an epitope of the *HPC2* protein. As used herein, "epitope" refers to an
30 antigenic determinant of a polypeptide. An epitope could comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids, and more usually consists of at least 8-10 such amino acids. Methods of determining the spatial conformation of such amino acids are known in the art.

For immunological purposes, tandem-repeat polypeptide segments may be used as immunogens, thereby producing highly antigenic proteins. Alternatively, such polypeptides will serve as highly efficient competitors for specific binding. Production of antibodies specific for HPC2 polypeptides or fragments thereof is described below.

5 The present invention also provides for fusion polypeptides, comprising HPC2 polypeptides and fragments. Homologous polypeptides may be fusions between two or more HPC2 polypeptide sequences or between the sequences of HPC2 and a related protein. Likewise, heterologous fusions may be constructed which would exhibit a combination of properties or activities of the derivative proteins. For example, ligand-binding or other domains
10 may be "swapped" between different new fusion polypeptides or fragments. Such homologous or heterologous fusion polypeptides may display, for example, altered strength or specificity of binding. Fusion partners include immunoglobulins, bacterial β -galactosidase, trpE, protein A, β -lactamase, alpha amylase, alcohol dehydrogenase and yeast alpha mating factor. See Godowski *et al.*, 1988.

15 Fusion proteins will typically be made by either recombinant nucleic acid methods, as described below, or may be chemically synthesized. Techniques for the synthesis of polypeptides are described, for example, in Merrifield, 1963.

"Protein purification" refers to various methods for the isolation of the HPC2 polypeptides from other biological material, such as from cells transformed with recombinant
20 nucleic acids encoding HPC2, and are well known in the art. For example, such polypeptides may be purified by immunoaffinity chromatography employing, e.g., the antibodies provided by the present invention. Various methods of protein purification are well known in the art, and include those described in Deutscher, 1990 and Scopes, 1982.

The terms "isolated", "substantially pure", and "substantially homogeneous" are used
25 interchangeably to describe a protein or polypeptide which has been separated from components which accompany it in its natural state. A monomeric protein is substantially pure when at least about 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure protein will typically comprise about 60 to 90% W/W of a protein sample, more usually about 95%, and preferably will be over about 99% pure. Protein purity or homogeneity may be
30 indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art which are utilized for purification.

An HPC2 protein is substantially free of naturally associated components when it is separated from the native contaminants which accompany it in its natural state. Thus, a polypeptide which is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be substantially free from its naturally associated components. A protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A polypeptide produced as an expression product of an isolated and manipulated genetic sequence is an "isolated polypeptide," as used herein, even if expressed in a homologous cell type. Synthetically made forms or molecules expressed by heterologous cells are inherently isolated molecules.

"Recombinant nucleic acid" is a nucleic acid which is not naturally occurring, or which is made by the artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions.

"Regulatory sequences" refers to those sequences normally within 100 kb of the coding region of a locus, but they may also be more distant from the coding region, which affect the expression of the gene (including transcription of the gene, and translation, splicing, stability or the like of the messenger RNA).

"Substantial homology or similarity". A nucleic acid or fragment thereof is "substantially homologous" ("or substantially similar") to another if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases.

Identity means the degree of sequence relatedness between two polypeptide or two polynucleotides sequences as determined by the identity of the match between two strings of such sequences. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Computational Molecular Biology, Lesk, A. M., ed., Oxford

University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribkov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Methods commonly employed to determine identity between two sequences include, but are not limited to those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H., and Lipman, D. (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Such methods are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux et al. (1984), BLASTP, BLASTN, FASTA (Altschul et al. (1990); Altschul et al. (1997)).

Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof will hybridize to another nucleic acid (or a complementary strand thereof) under selective hybridization conditions, to a strand, or to its complement. Selectivity of hybridization exists when hybridization which is substantially more selective than total lack of specificity occurs. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 14 nucleotides, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90%. See, Kanehisa, 1984. The length of homology comparison, as described, may be over longer stretches, and in certain embodiments will often be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions will generally include temperatures in excess of 30°C, typically in excess of 37°C, and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. However, the combination of parameters is much more important than the measure of any single parameter. See, e.g., Wetmur and Davidson, 1968.

Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

The terms "**substantial homology**" or "**substantial identity**", when referring to polypeptides, indicate that the polypeptide or protein in question exhibits at least about 30% identity with an entire naturally-occurring protein or a portion thereof, usually at least about 70% identity, more usually at least about 80% identity, preferably at least about 90% identity, and more preferably at least about 95% identity.

Homology, for polypeptides, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705, as well as the software described above with reference to nucleic acid homology. Protein analysis software matches similar sequences using measures of homology assigned to various substitutions, deletions and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

"**Substantially similar function**" refers to the function of a modified nucleic acid or a modified protein, with reference to the wild-type HPC2 nucleic acid or wild-type HPC2 polypeptide. The modified polypeptide will be substantially homologous to the wild-type HPC2 polypeptide and will have substantially the same function. The modified polypeptide may have an altered amino acid sequence and/or may contain modified amino acids. In addition to the similarity of function, the modified polypeptide may have other useful properties, such as a longer half-life. The similarity of function (activity) of the modified polypeptide may be substantially the same as the activity of the wild-type HPC2 polypeptide. Alternatively, the similarity of function (activity) of the modified polypeptide may be higher than the activity of the wild-type HPC2 polypeptide. The modified polypeptide is synthesized using conventional techniques, or is encoded by a modified nucleic acid and produced using conventional techniques. The modified nucleic acid is prepared by conventional techniques. A nucleic acid with a function substantially similar to the wild-type HPC2 gene function produces the modified protein described above.

A polypeptide "**fragment**," "**portion**" or "**segment**" is a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous

amino acids, typically at least about nine to 13 contiguous amino acids and, most preferably, at least about 20 to 30 or more contiguous amino acids.

The polypeptides of the present invention, if soluble, may be coupled to a solid-phase support, e.g., nitrocellulose, nylon, column packing materials (e.g., Sepharose beads), magnetic
5 beads, glass wool, plastic, metal, polymer gels, cells, or other substrates. Such supports may take the form, for example, of beads, wells, dipsticks, or membranes.

"**Target region**" refers to a region of the nucleic acid which is amplified and/or detected. The term "**target sequence**" refers to a sequence with which a probe or primer will form a stable hybrid under desired conditions.

10 The practice of the present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, genetics, and immunology. See, e.g., Maniatis *et al.*, 1982; Sambrook *et al.*, 1989; Ausubel *et al.*, 1992; Glover, 1985; Anand, 1992; Guthrie and Fink, 1991. A general discussion of techniques and materials for human gene mapping, including mapping of human chromosome 1, is provided,
15 e.g., in White and Lalouel, 1988.

Preparation of recombinant or chemically
synthesized nucleic acids; vectors, transformation, host cells

Large amounts of the polynucleotides of the present invention may be produced by replication in a suitable host cell. Natural or synthetic polynucleotide fragments coding for a
20 desired fragment will be incorporated into recombinant polynucleotide constructs, usually DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the polynucleotide constructs will be suitable for replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to (with and without integration within the genome) cultured mammalian or plant or other eukaryotic cell lines. The purification
25 of nucleic acids produced by the methods of the present invention is described, e.g., in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g., by the phosphoramidite method described by Beaucage and Caruthers, 1981 or the triester method according to Matteucci and Caruthers, 1981, and may be performed on
30 commercial, automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single-stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strands together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Polynucleotide constructs prepared for introduction into a prokaryotic or eukaryotic host may comprise a replication system recognized by the host, including the intended polynucleotide fragment encoding the desired polypeptide, and will preferably also include transcription and translational initiation regulatory sequences operably linked to the polypeptide encoding segment. Expression vectors may include, for example, an origin of replication or autonomously replicating sequence (ARS) and expression control sequences, a promoter, an enhancer and necessary processing information sites, such as ribosome-binding sites, RNA splice sites, polyadenylation sites, transcriptional terminator sequences, and mRNA stabilizing sequences. Secretion signals may also be included where appropriate, whether from a native HPC2 protein or from other receptors or from secreted polypeptides of the same or related species, which allow the protein to cross and/or lodge in cell membranes, and thus attain its functional topology, or be secreted from the cell. Such vectors may be prepared by means of standard recombinant techniques well known in the art and discussed, for example, in Sambrook *et al.*, 1989 or Ausubel *et al.* 1992.

An appropriate promoter and other necessary vector sequences will be selected so as to be functional in the host, and may include, when appropriate, those naturally associated with HPC2 genes. Examples of workable combinations of cell lines and expression vectors are described in Sambrook *et al.*, 1989 or Ausubel *et al.*, 1992; see also, e.g., Metzger *et al.*, 1988. Many useful vectors are known in the art and may be obtained from such vendors as Stratagene, New England BioLabs, Promega Biotech, and others. Promoters such as the trp, lac and phage promoters, tRNA promoters and glycolytic enzyme promoters may be used in prokaryotic hosts. Useful yeast promoters include promoter regions for metallothionein, 3-phosphoglycerate kinase or other glycolytic enzymes such as enolase or glyceraldehyde-3-phosphate dehydrogenase, enzymes responsible for maltose and galactose utilization, and others. Vectors and promoters suitable for use in yeast expression are further described in Hitzeman *et al.*, EP 73,675A. Appropriate non-native mammalian promoters might include the early and late promoters from SV40 (Fiers *et al.*, 1978) or promoters derived from murine Moloney leukemia virus, mouse tumor virus, avian sarcoma viruses, adenovirus II, bovine papilloma virus or polyoma. Infection promoters may be derived from baculovirus. In addition, the construct may be joined to an amplifiable gene (e.g., DHFR) so that multiple copies of the gene may be made. For appropriate enhancer and other expression control sequences, see also *Enhancers and Eukaryotic Gene Expression*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1983). See also, e.g., U.S. Patent Nos. 5,691,198; 5,735,500; 5,747,469 and 5,436,146.

While such expression vectors may replicate autonomously, they may also replicate by being inserted into the genome of the host cell, by methods well known in the art.

Expression and cloning vectors will likely contain a selectable marker, a gene encoding a protein necessary for survival or growth of a host cell transformed with the vector. The presence of this gene ensures growth of only those host cells which express the inserts. Typical selection genes encode proteins that a) confer resistance to antibiotics or other toxic substances, e.g. ampicillin, neomycin, methotrexate, etc.; b) complement auxotrophic deficiencies, or c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. The choice of the proper selectable marker will depend on the host cell, and appropriate markers for different hosts are well known in the art.

The vectors containing the nucleic acids of interest can be transcribed *in vitro*, and the resulting RNA introduced into the host cell by well-known methods, e.g., by injection (see, Kubo *et al.*, 1988), or the vectors can be introduced directly into host cells by methods well known in the art, which vary depending on the type of cellular host, including electroporation; transfection employing calcium chloride, rubidium chloride, calcium phosphate, DEAE-dextran, or other substances; microprojectile bombardment; lipofection; infection (where the vector is an infectious agent, such as a retroviral genome); and other methods. See generally, Sambrook *et al.*, 1989 and Ausubel *et al.*, 1992. The introduction of the polynucleotides into the host cell by any method known in the art, including, *inter alia*, those described above, will be referred to herein as "transformation." The cells into which have been introduced nucleic acids described above are meant to also include the progeny of such cells.

Large quantities of the nucleic acids and polypeptides of the present invention may be prepared by expressing the HPC2 nucleic acids or portions thereof in vectors or other expression vehicles in compatible prokaryotic or eukaryotic host cells. The most commonly used prokaryotic hosts are strains of *Escherichia coli*, although other prokaryotes, such as *Bacillus subtilis* or *Pseudomonas* may also be used.

Mammalian or other eukaryotic host cells, such as those of yeast, filamentous fungi, plant, insect, or amphibian or avian species, may also be useful for production of the proteins of the present invention. Propagation of mammalian cells in culture is *per se* well known. See, Jakoby and Pastan, 1979. Examples of commonly used mammalian host cell lines are VERO and HeLa cells, Chinese hamster ovary (CHO) cells, and WI38, BHK, and COS cell lines, although it will be appreciated by the skilled practitioner that other cell lines may be appropriate,

e.g., to provide higher expression, desirable glycosylation patterns, or other features. An example of a commonly used insect cell line is SF9.

Clones are selected by using markers depending on the mode of the vector construction. The marker may be on the same or a different DNA molecule, preferably the same DNA molecule. In prokaryotic hosts, the transformant may be selected, e.g., by resistance to ampicillin, tetracycline or other antibiotics. Production of a particular product based on temperature sensitivity may also serve as an appropriate marker.

Prokaryotic or eukaryotic cells transformed with the polynucleotides of the present invention will be useful not only for the production of the nucleic acids and polypeptides of the present invention, but also, for example, in studying the characteristics of HPC2 polypeptides.

The HPC2 gene products can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, guinea pigs, pigs, micro-pigs, goats and non-human primates, e.g., baboons, monkeys and chimpanzees, may be used to generate HPC2 transgenic animals.

Any technique known in the art may be used to introduce the HPC2 gene transgene into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (U.S. Patent No. 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., 1985); gene targeting in embryonic stem cells (Thompson et al., 1989); electroporation of embryos (Lo, 1983); and sperm-mediated gene transfer (Lavitrano et al., 1989); etc. For a review of such techniques, see Gordon (1989), which is incorporated by reference herein in its entirety.

The present invention provides for transgenic animals that carry the HPC2 transgene in all their cells, as well as animals which carry the transgene in some, but not all of their cells, i.e., mosaic animals. The transgene may be integrated as a single transgene or in concatamers, e.c., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the HPC2 gene transgene be integrated into the chromosomal site of the endogenous HPC2 gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous HPC2 gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the

endogenous HPC2 gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous HPC2 gene in only that cell type, by following, for example, the teaching of Gu et al. (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant HPC2 gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to assay whether integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and RT-PCR. Samples of HPC2 gene-expressing tissue, may also be evaluated immunocytochemically using antibodies specific for the HPC2 transgene product.

Antisense polynucleotide sequences are useful in preventing or diminishing the expression of the HPC2 locus, as will be appreciated by those skilled in the art. For example, polynucleotide vectors containing all or a portion of the HPC2 locus or other sequences from the HPC2 region (particularly those flanking the HPC2 locus) may be placed under the control of a promoter in an antisense orientation and introduced into a cell. Expression of such an antisense construct within a cell will interfere with HPC2 transcription and/or translation and/or replication.

The probes and primers based on the HPC2 gene sequences disclosed herein are used to identify homologous HPC2 gene sequences and proteins in other species. These HPC2 gene sequences and proteins are used in the diagnostic/prognostic, therapeutic and drug screening methods described herein for the species from which they have been isolated.

Methods of Use: Nucleic Acid Diagnosis and Diagnostic Kits

In order to detect the presence of an HPC2 allele predisposing an individual to cancer, a biological sample such as blood is prepared and analyzed for the presence or absence of susceptibility alleles of HPC2. In order to detect the presence of neoplasia, the progression toward malignancy of a precursor lesion, or as a prognostic indicator, a biological sample of the lesion is prepared and analyzed for the presence or absence of mutant alleles of HPC2. Results of these tests and interpretive information are returned to the health care provider for communication to the tested individual. Such diagnoses may be performed by diagnostic

laboratories, or, alternatively, diagnostic kits are manufactured and sold to health care providers or to private individuals for self-diagnosis.

Initially, the screening method involves amplification of the relevant HPC2 sequences. In another preferred embodiment of the invention, the screening method involves a non-PCR based strategy. Such screening methods include two-step label amplification methodologies that are well known in the art. Both PCR and non-PCR based screening strategies can detect target sequences with a high level of sensitivity.

The most popular method used today is target amplification. Here, the target nucleic acid sequence is amplified with polymerases. One particularly preferred method using polymerase-driven amplification is the polymerase chain reaction (PCR). The polymerase chain reaction and other polymerase-driven amplification assays can achieve over a million-fold increase in copy number through the use of polymerase-driven amplification cycles. Once amplified, the resulting nucleic acid can be sequenced or used as a substrate for DNA probes.

When the probes are used to detect the presence of the target sequences (for example, in screening for cancer susceptibility), the biological sample to be analyzed, such as blood or serum, may be treated, if desired, to extract the nucleic acids. The sample nucleic acid may be prepared in various ways to facilitate detection of the target sequence; e.g. denaturation, restriction digestion, electrophoresis or dot blotting. The targeted region of the analyte nucleic acid usually must be at least partially single-stranded to form hybrids with the targeting sequence of the probe. If the sequence is naturally single-stranded, denaturation will not be required. However, if the sequence is double-stranded, the sequence will probably need to be denatured. Denaturation can be carried out by various techniques known in the art.

Analyte nucleic acid and probe are incubated under conditions which promote stable hybrid formation of the target sequence in the probe with the putative targeted sequence in the analyte. The region of the probes which is used to bind to the analyte can be made completely complementary to the targeted region of human chromosome 17. Therefore, high stringency conditions are desirable in order to prevent false positives. However, conditions of high stringency are used only if the probes are complementary to regions of the chromosome which are unique in the genome. The stringency of hybridization is determined by a number of factors during hybridization and during the washing procedure, including temperature, ionic strength, base composition, probe length, and concentration of formamide. These factors are outlined in, for example, Maniatis *et al.*, 1982 and Sambrook *et al.*, 1989. Under certain circumstances, the

formation of higher order hybrids, such as triplexes, quadraplexes, etc., may be desired to provide the means of detecting target sequences.

Detection, if any, of the resulting hybrid is usually accomplished by the use of labeled probes. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labeled, either directly or indirectly. Suitable labels, and methods for labeling probes and ligands are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (e.g., nick translation, random priming or kinasing), biotin, fluorescent groups, chemiluminescent groups (e.g., dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, gold nanoparticles and the like. Variations of this basic scheme are known in the art, and include those variations that facilitate separation of the hybrids to be detected from extraneous materials and/or that amplify the signal from the labeled moiety. A number of these variations are reviewed in, e.g., Matthews and Kricka, 1988; Landegren *et al.*, 1988; Mifflin, 1989; U.S. Patent 4,868,105, and in EPO Publication No. 225,807.

As noted above, non-PCR based screening assays are also contemplated in this invention. This procedure hybridizes a nucleic acid probe (or an analog such as a methyl phosphonate backbone replacing the normal phosphodiester), to the low level DNA target. This probe may have an enzyme covalently linked to the probe, such that the covalent linkage does not interfere with the specificity of the hybridization. This enzyme-probe-conjugate-target nucleic acid complex can then be isolated away from the free probe enzyme conjugate and a substrate is added for enzyme detection. Enzymatic activity is observed as a change in color development or luminescent output resulting in a 10^3 - 10^6 increase in sensitivity. For an example relating to the preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes see Jablonski *et al.*, 1986.

Two-step label amplification methodologies are known in the art. These assays work on the principle that a small ligand (such as digoxigenin, biotin, or the like) is attached to a nucleic acid probe capable of specifically binding HPC2. Allele specific probes are also contemplated within the scope of this example and exemplary allele specific probes include probes encompassing the predisposing or potentially predisposing mutations summarized in Table 8 of this patent application.

In one example, the small ligand attached to the nucleic acid probe is specifically recognized by an antibody-enzyme conjugate. In one embodiment of this example, digoxigenin is attached to the nucleic acid probe. Hybridization is detected by an antibody-alkaline

phosphatase conjugate which turns over a chemiluminescent substrate. For methods for labeling nucleic acid probes according to this embodiment see Martin *et al.*, 1990. In a second example, the small ligand is recognized by a second ligand-enzyme conjugate that is capable of specifically complexing to the first ligand. A well known embodiment of this example is the biotin-avidin type of interactions. For methods for labeling nucleic acid probes and their use in biotin-avidin based assays see Rigby *et al.*, 1977 and Nguyen *et al.*, 1992.

It is also contemplated within the scope of this invention that the nucleic acid probe assays of this invention will employ a cocktail of nucleic acid probes capable of detecting HPC2. Thus, in one example to detect the presence of HPC2 in a cell sample, more than one probe complementary to HPC2 is employed and in particular the number of different probes is alternatively 2, 3, or 5 different nucleic acid probe sequences. In another example, to detect the presence of mutations in the HPC2 gene sequence in a patient, more than one probe complementary to HPC2 is employed where the cocktail includes probes capable of binding to the allele-specific mutations identified in populations of patients with alterations in HPC2. In this embodiment, any number of probes can be used, and will preferably include probes corresponding to the major gene mutations identified as predisposing an individual to prostate cancer.

Methods of Use: Peptide Diagnosis and Diagnostic Kits

The neoplastic condition of lesions can also be detected on the basis of the alteration of wild-type HPC2 polypeptide. Such alterations can be determined by sequence analysis in accordance with conventional techniques. More preferably, antibodies (polyclonal or monoclonal) are used to detect differences in, or the absence of, HPC2 peptides. The antibodies may be prepared as discussed above under the heading "Antibodies" and as further shown in Examples 9 and 10. Other techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparations claimed in this invention. In a preferred embodiment of the invention, antibodies will immunoprecipitate HPC2 proteins from solution as well as react with HPC2 protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, antibodies will detect HPC2 proteins in paraffin or frozen tissue sections, using immunocytochemical techniques.

Preferred embodiments relating to methods for detecting HPC2 or its mutations include enzyme linked immunosorbent assays (ELISA), radioimmunoassays (RIA), immunoradiometric assays (IRMA) and immunoenzymatic assays (IEMA), including sandwich assays using monoclonal and/or polyclonal antibodies. Exemplary sandwich assays are described by David *et*

al. in U.S. Patent Nos. 4,376,110 and 4,486,530, hereby incorporated by reference, and exemplified in Example 12.

Methods of Use: Drug Screening

This invention is particularly useful for screening compounds by using the HPC2 polypeptide or binding fragment thereof in any of a variety of drug screening techniques.

The HPC2 polypeptide or fragment employed in such a test may either be free in solution, affixed to a solid support, or borne on a cell surface. One method of drug screening utilizes eucaryotic or procaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, for the formation of complexes between an HPC2 polypeptide or fragment and the agent being tested, or examine the degree to which the formation of a complex between an HPC2 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Thus, the present invention provides methods of screening for drugs comprising contacting such an agent with an HPC2 polypeptide or fragment thereof and assaying (i) for the presence of a complex between the agent and the HPC2 polypeptide or fragment, or (ii) for the presence of a complex between the HPC2 polypeptide or fragment and a ligand, by methods well known in the art. In such competitive binding assays the HPC2 polypeptide or fragment is typically labeled. Free HPC2 polypeptide or fragment is separated from that present in a protein:protein complex, and the amount of free (i.e., uncomplexed) label is a measure of the binding of the agent being tested to HPC2 or its interference with HPC2:ligand binding, respectively. One may also measure the amount of bound, rather than free, HPC2. It is also possible to label the ligand rather than the HPC2 and to measure the amount of ligand binding to HPC2 in the presence and in the absence of the drug being tested.

Another technique for drug screening provides high throughput screening for compounds having suitable binding affinity to the HPC2 polypeptides and is described in detail in Geysen (published PCT WO 84/03564). Briefly stated, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with HPC2 polypeptide and washed. Bound HPC2 polypeptide is then detected by methods well known in the art.

Purified HPC2 can be coated directly onto plates for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the HPC2 polypeptide on the solid phase.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the HPC2 polypeptide compete with a test compound for binding to the HPC2 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the HPC2 polypeptide.

A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional HPC2 gene. These host cell lines or cells are defective at the HPC2 polypeptide level. The host cell lines or cells are grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of HPC2 defective cells.

Briefly, a method of screening for a substance which modulates activity of a polypeptide may include contacting one or more test substances with the polypeptide in a suitable reaction medium, testing the activity of the treated polypeptide and comparing that activity with the activity of the polypeptide in comparable reaction medium untreated with the test substance or substances. A difference in activity between the treated and untreated polypeptides is indicative of a modulating effect of the relevant test substance or substances.

Prior to or as well as being screened for modulation of activity, test substances may be screened for ability to interact with the polypeptide, e.g., in a yeast two-hybrid system (e.g., Bartel et al., 1993; Fields and Song, 1989; Chevray and Nathans, 1992; Lee et al., 1995). This system may be used as a coarse screen prior to testing a substance for actual ability to modulate activity of the polypeptide. Alternatively, the screen could be used to screen test substances for binding to an HPC2 specific binding partner, or to find mimetics of an HPC2 polypeptide.

Methods of Use: Rational Drug Design

The goal of rational drug design is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which, e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, 1991. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., HPC2 polypeptide) or, for example, of the HPC2-receptor or ligand complex, by x-ray crystallography, by computer modeling or most typically, by a

combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson *et al.*, 1990). In addition, peptides (e.g., HPC2 polypeptide) are analyzed by an alanine scan (Wells, 1991). In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original receptor. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved HPC2 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of HPC2 polypeptide activity. By virtue of the availability of cloned HPC2 sequences, sufficient amounts of the HPC2 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the HPC2 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

Following identification of a substance which modulates or affects polypeptide activity, the substance may be investigated further. Furthermore, it may be manufactured and/or used in preparation, i.e., manufacture or formulation, or a composition such as a medicament, pharmaceutical composition or drug. These may be administered to individuals.

Thus, the present invention extends in various aspects not only to a substance identified using a nucleic acid molecule as a modulator of polypeptide activity, in accordance with what is disclosed herein, but also a pharmaceutical composition, medicament, drug or other composition comprising such a substance, a method comprising administration of such a composition comprising such a substance, a method comprising administration of such a composition to a patient, e.g., for treatment of prostate cancer, use of such a substance in the manufacture of a composition for administration, e.g., for treatment of prostate cancer, and a method of making a

pharmaceutical composition comprising admixing such a substance with a pharmaceutically acceptable excipient, vehicle or carrier, and optionally other ingredients.

A substance identified as a modulator of polypeptide function may be peptide or non-peptide in nature. Non-peptide “small molecules” are often preferred for many *in vivo* pharmaceutical uses. Accordingly, a mimetic or mimic of the substance (particularly if a peptide) may be designed for pharmaceutical use.

The designing of mimetics to a known pharmaceutically active compound is a known approach to the development of pharmaceuticals based on a “lead” compound. This might be desirable where the active compound is difficult or expensive to synthesize or where it is unsuitable for a particular method of administration, e.g., pure peptides are unsuitable active agents for oral compositions as they tend to be quickly degraded by proteases in the alimentary canal. Mimetic design, synthesis and testing is generally used to avoid randomly screening large numbers of molecules for a target property.

There are several steps commonly taken in the design of a mimetic from a compound having a given target property. First, the particular parts of the compound that are critical and/or important in determining the target property are determined. In the case of a peptide, this can be done by systematically varying the amino acid residues in the peptide, e.g., by substituting each residue in turn. Alanine scans of peptide are commonly used to refine such peptide motifs. These parts or residues constituting the active region of the compound are known as its “pharmacophore”.

Once the pharmacophore has been found, its structure is modeled according to its physical properties, e.g., stereochemistry, bonding, size and/or charge, using data from a range of sources, e.g., spectroscopic techniques, x-ray diffraction data and NMR. Computational analysis, similarity mapping (which models the charge and/or volume of a pharmacophore, rather than the bonding between atoms) and other techniques can be used in this modeling process.

In a variant of this approach, the three-dimensional structure of the ligand and its binding partner are modeled. This can be especially useful where the ligand and/or binding partner change conformation on binding, allowing the model to take account of this in the design of the mimetic.

A template molecule is then selected onto which chemical groups which mimic the pharmacophore can be grafted. The template molecule and the chemical groups grafted onto it can conveniently be selected so that the mimetic is easy to synthesize, is likely to be

pharmacologically acceptable, and does not degrade *in vivo*, while retaining the biological activity of the lead compound. Alternatively, where the mimetic is peptide-based, further stability can be achieved by cyclizing the peptide, increasing its rigidity. The mimetic or mimetics found by this approach can then be screened to see whether they have the target
 5 property, or to what extent they exhibit it. Further optimization or modification can then be carried out to arrive at one or more final mimetics for *in vivo* or clinical testing.

Methods of Use: Gene Therapy

According to the present invention, a method is also provided of supplying wild-type HPC2 function to a cell which carries mutant HPC2 alleles. Supplying such a function should
 10 suppress neoplastic growth of the recipient cells. The wild-type HPC2 gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene fragment is introduced and expressed in a cell carrying a mutant HPC2 allele, the gene fragment should encode a part of the HPC2 protein which is required for non-neoplastic growth
 15 of the cell. More preferred is the situation where the wild-type HPC2 gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant HPC2 gene present in the cell. Such recombination requires a double recombination event which results in the correction of the HPC2 gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any
 20 suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate coprecipitation and viral transduction are known in the art, and the choice of method is within the competence of the practitioner. Cells transformed with the wild-type HPC2 gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the HPC2 gene or fragment, where applicable, may be
 25 employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in both cancerous and pre-cancerous cells, in which the level of HPC2 polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression
 30 of a given HPC2 gene even in those tumor cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example, as described by Friedman (1991) or Culver (1996). Cells from a patient's tumor would

be first analyzed by the diagnostic methods described above, to ascertain the production of HPC2 polypeptide in the tumor cells. A virus or plasmid vector (see further details below), containing a copy of the HPC2 gene linked to expression control elements and capable of replicating inside the tumor cells, is prepared. Alternatively, the vector may be replication deficient and is replicated in helper cells for use in gene therapy. Suitable vectors are known, such as disclosed in U.S. Patent 5,252,479 and PCT published application WO 93/07282 and U.S. Patent Nos. 5,691,198; 5,747,469; 5,436,146 and 5,753,500. The vector is then injected into the patient, either locally at the site of the tumor or systemically (in order to reach any tumor cells that may have metastasized to other sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted tumor cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods of the present invention. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g., SV40 (Madzak *et al.*, 1992), adenovirus (Berkner, 1992; Berkner *et al.*, 1988; Gorziglia and Kapikian, 1992; Quantin *et al.*, 1992; Rosenfeld *et al.*, 1992; Wilkinson and Akrigg, 1992; Stratford-Perricaudet *et al.*, 1990; Schneider *et al.*, 1998), vaccinia virus (Moss, 1992; Moss, 1996), adeno-associated virus (Muzyczka, 1992; Ohi *et al.*, 1990; Russell and Hirata, 1998), herpes viruses including HSV and EBV (Margolskee, 1992; Johnson *et al.*, 1992; Fink *et al.*, 1992; Breakefield and Geller, 1987; Freese *et al.*, 1990; Fink *et al.*, 1996), lentiviruses (Naldini *et al.*, 1996), Sindbis and Semliki Forest virus (Berglund *et al.*, 1993), and retroviruses of avian (Bandyopadhyay and Temin, 1984; Petropoulos *et al.*, 1992), murine (Miller, 1992; Miller *et al.*, 1985; Sorge *et al.*, 1984; Mann and Baltimore, 1985; Miller *et al.*, 1988), and human origin (Shimada *et al.*, 1991; Helseth *et al.*, 1990; Page *et al.*, 1990; Buchschacher and Panganiban, 1992). Most human gene therapy protocols have been based on disabled murine retroviruses, although adenovirus and adeno-associated virus are also being used.

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Graham and van der Eb, 1973; Pellicer *et al.*, 1980); mechanical techniques, for example microinjection (Anderson *et al.*, 1980; Gordon *et al.*, 1980; Brinster *et al.*, 1981; Costantini and Lacy, 1981); membrane fusion-mediated transfer via liposomes (Felgner *et al.*, 1987; Wang and Huang, 1989; Kaneda *et al.*, 1989; Stewart *et al.*, 1992; Nabel *et al.*, 1990; Lim *et al.*, 1991); and direct DNA uptake and receptor-mediated DNA transfer (Wolff *et al.*, 1990; Wu *et al.*, 1991; Zenke *et al.*, 1990; Wu *et al.*, 1989; Wolff *et al.*,

1991; Wagner *et al.*, 1990; Wagner *et al.*, 1991; Cotten *et al.*, 1990; Curiel *et al.*, 1991; Curiel *et al.*, 1992). Viral-mediated gene transfer can be combined with direct *in vivo* gene transfer using liposome delivery, allowing one to direct the viral vectors to the tumor cells and not into the surrounding nondividing cells. Alternatively, the retroviral vector producer cell line can be injected into tumors (Culver *et al.*, 1992). Injection of producer cells would then provide a continuous source of vector particles. This technique has been approved for use in humans with inoperable brain tumors.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged. For other techniques for the delivery of adenovirus based vectors see Schneider *et al.* (1998) and U.S. Patent Nos. 5,691,198; 5,747,469; 5,436,146 and 5,753,500.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumor deposits, for example, following direct *in situ* administration (Nabel, 1992).

Expression vectors in the context of gene therapy are meant to include those constructs containing sequences sufficient to express a polynucleotide that has been cloned therein. In viral expression vectors, the construct contains viral sequences sufficient to support packaging of the construct. If the polynucleotide encodes *HPC2*, expression will produce HPC2. If the polynucleotide encodes an antisense polynucleotide or a ribozyme, expression will produce the antisense polynucleotide or ribozyme. Thus in this context, expression does not require that a protein product be synthesized. In addition to the polynucleotide cloned into the expression vector, the vector also contains a promoter functional in eukaryotic cells. The cloned polynucleotide sequence is under control of this promoter. Suitable eukaryotic promoters include those described above. The expression vector may also include sequences, such as selectable markers and other sequences described herein.

Gene transfer techniques which target DNA directly to prostate tissues, e.g., epithelial cells of the prostate, are preferred. Receptor-mediated gene transfer, for example, is accomplished by the conjugation of DNA (usually in the form of covalently closed supercoiled plasmid) to a protein ligand via polylysine. Ligands are chosen on the basis of the presence of

the corresponding ligand receptors on the cell surface of the target cell/tissue type. One appropriate receptor/ligand pair may include the estrogen receptor and its ligand, estrogen (and estrogen analogues). These ligand-DNA conjugates can be injected directly into the blood if desired and are directed to the target tissue where receptor binding and internalization of the DNA-protein complex occurs. To overcome the problem of intracellular destruction of DNA, coinfection with adenovirus can be included to disrupt endosome function.

The therapy involves two steps which can be performed singly or jointly. In the first step, prepubescent females who carry an HPC2 susceptibility allele are treated with a gene delivery vehicle such that some or all of their mammary ductal epithelial precursor cells receive at least one additional copy of a functional normal HPC2 allele. In this step, the treated individuals have reduced risk of prostate cancer to the extent that the effect of the susceptible allele has been countered by the presence of the normal allele. In the second step of a preventive therapy, predisposed young females, in particular women who have received the proposed gene therapeutic treatment, undergo hormonal therapy to mimic the effects on the prostate of a full term pregnancy.

Methods of Use: Peptide Therapy

Peptides which have HPC2 activity can be supplied to cells which carry mutant or missing HPC2 alleles. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, HPC2 polypeptide can be extracted from HPC2-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize HPC2 protein. Any of such techniques can provide the preparation of the present invention which comprises the HPC2 protein. Preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or *in vitro*.

Active HPC2 molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the HPC2 gene product may be sufficient to affect tumor growth. Supply of molecules with HPC2 activity should lead to partial reversal of the neoplastic state. Other molecules with HPC2 activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function are also used for peptide therapy.

Methods of Use: Transformed Hosts

Similarly, cells and animals which carry a mutant HPC2 allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with HPC2 mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the HPC2 allele, as described above. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant HPC2 alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. Alternatively, the endogenous HPC2 gene(s) of the animals may be disrupted by insertion or deletion mutation or other genetic alterations using conventional techniques (Capecci, 1989; Valancius and Smithies, 1991; Hasty *et al.*, 1991; Shinkai *et al.*, 1992; Mombaerts *et al.*, 1992; Philpott *et al.*, 1992; Snouwaert *et al.*, 1992; Donehower *et al.*, 1992) to produce knockout or transplacement animals. A transplacement is similar to a knockout because the endogenous gene is replaced, but in the case of a transplacement the replacement is by another version of the same gene. After test substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of the cancers identified herein. These animal models provide an extremely important testing vehicle for potential therapeutic products.

In one embodiment of the invention, transgenic animals are produced which contain a functional transgene encoding a functional HPC2 polypeptide or variants thereof. Transgenic animals expressing *HPC2* transgenes, recombinant cell lines derived from such animals and transgenic embryos may be useful in methods for screening for and identifying agents that induce or repress function of HPC2. Transgenic animals of the present invention also can be used as models for studying indications such as disease.

In one embodiment of the invention, a *HPC2* transgene is introduced into a non-human host to produce a transgenic animal expressing a human or murine *HPC2* gene. The transgenic animal is produced by the integration of the transgene into the genome in a manner that permits

the expression of the transgene. Methods for producing transgenic animals are generally described by Wagner and Hoppe (U.S. Patent No. 4,873,191; which is incorporated herein by reference), Brinster *et al.* 1985; which is incorporated herein by reference in its entirety) and in "Manipulating the Mouse Embryo; A Laboratory Manual" 2nd edition (eds., Hogan, Beddington, Costantini and Long, Cold Spring Harbor Laboratory Press, 1994; which is incorporated herein by reference in its entirety).

It may be desirable to replace the endogenous *HPC2* by homologous recombination between the transgene and the endogenous gene; or the endogenous gene may be eliminated by deletion as in the preparation of "knock-out" animals. Typically, a *HPC2* gene flanked by genomic sequences is transferred by microinjection into a fertilized egg. The microinjected eggs are implanted into a host female, and the progeny are screened for the expression of the transgene. Transgenic animals may be produced from the fertilized eggs from a number of animals including, but not limited to reptiles, amphibians, birds, mammals, and fish. Within a particularly preferred embodiment, transgenic mice are generated which overexpress *HPC2* or express a mutant form of the polypeptide. Alternatively, the absence of a *HPC2* in "knock-out" mice permits the study of the effects that loss of *HPC2* protein has on a cell *in vivo*. Knock-out mice also provide a model for the development of *HPC2*-related cancers.

Methods for producing knockout animals are generally described by Shastry (1995, 1998) and Osterrieder and Wolf (1998). The production of conditional knockout animals, in which the gene is active until knocked out at the desired time is generally described by Feil *et al.* (1996), Gagneten *et al.* (1997) and Lobe and Nagy (1998). Each of these references is incorporated herein by reference.

As noted above, transgenic animals and cell lines derived from such animals may find use in certain testing experiments. In this regard, transgenic animals and cell lines capable of expressing wild-type or mutant *HPC2* may be exposed to test substances. These test substances can be screened for the ability to reduce overexpression of wild-type *HPC2* or impair the expression or function of mutant *HPC2*.

Pharmaceutical Compositions and Routes of Administration

The *HPC2* polypeptides, antibodies, peptides and nucleic acids of the present invention can be formulated in pharmaceutical compositions, which are prepared according to conventional pharmaceutical compounding techniques. See, for example, Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA). The composition may contain the active agent or pharmaceutically acceptable salts of the active agent. These compositions may

comprise, in addition to one of the active substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The carrier may take a wide variety of forms depending on the form of preparation desired for administration, e.g.,

5 intravenous, oral, intrathecal, epineural or parenteral.

For oral administration, the compounds can be formulated into solid or liquid preparations such as capsules, pills, tablets, lozenges, melts, powders, suspensions or emulsions. In preparing the compositions in oral dosage form, any of the usual pharmaceutical media may be employed, such as, for example, water, glycols, oils, alcohols, flavoring agents, preservatives,

10 coloring agents, suspending agents, and the like in the case of oral liquid preparations (such as, for example, suspensions, elixirs and solutions); or carriers such as starches, sugars, diluents, granulating agents, lubricants, binders, disintegrating agents and the like in the case of oral solid preparations (such as, for example, powders, capsules and tablets). Because of their ease in administration, tablets and capsules represent the most advantageous oral dosage unit form, in

15 which case solid pharmaceutical carriers are obviously employed. If desired, tablets may be sugar-coated or enteric-coated by standard techniques. The active agent can be encapsulated to make it stable to passage through the gastrointestinal tract while at the same time allowing for passage across the blood brain barrier. See for example, WO 96/11698.

For parenteral administration, the compound may be dissolved in a pharmaceutical

20 carrier and administered as either a solution or a suspension. Illustrative of suitable carriers are water, saline, dextrose solutions, fructose solutions, ethanol, or oils of animal, vegetative or synthetic origin. The carrier may also contain other ingredients, for example, preservatives, suspending agents, solubilizing agents, buffers and the like. When the compounds are being administered intrathecally, they may also be dissolved in cerebrospinal fluid.

The active agent is preferably administered in a therapeutically effective amount. The actual amount administered, and the rate and time-course of administration, will depend on the nature and severity of the condition being treated. Prescription of treatment, e.g. decisions on dosage, timing, etc., is within the responsibility of general practitioners or specialists, and typically takes account of the disorder to be treated, the condition of the individual patient, the

30 site of delivery, the method of administration and other factors known to practitioners. Examples of techniques and protocols can be found in *Remington's Pharmaceutical Sciences*.

Alternatively, targeting therapies may be used to deliver the active agent more specifically to certain types of cell, by the use of targeting systems such as antibodies or cell

specific ligands. Targeting may be desirable for a variety of reasons, e.g. if the agent is unacceptably toxic, or if it would otherwise require too high a dosage, or if it would not otherwise be able to enter the target cells.

Instead of administering these agents directly, they could be produced in the target cell, e.g. in a viral vector such as described above or in a cell based delivery system such as described in U.S. Patent No. 5,550,050 and published PCT application Nos. WO 92/19195, WO 94/25503, WO 95/01203, WO 95/05452, WO 96/02286, WO 96/02646, WO 96/40871, WO 96/40959 and WO 97/12635, designed for implantation in a patient. The vector could be targeted to the specific cells to be treated, or it could contain regulatory elements which are more tissue specific to the target cells. The cell based delivery system is designed to be implanted in a patient's body at the desired target site and contains a coding sequence for the active agent. Alternatively, the agent could be administered in a precursor form for conversion to the active form by an activating agent produced in, or targeted to, the cells to be treated. See for example, EP 425,731A and WO 90/07936.

The identification of the association between the HPC2 gene mutations prostate cancer permits the early presymptomatic screening of individuals to identify those at risk for developing prostate cancer. To identify such individuals, HPC2 alleles are screened for mutations either directly or after cloning the alleles. The alleles are tested for the presence of nucleic acid sequence differences from the normal allele using any suitable technique, including but not limited to, one of the following methods: fluorescent *in situ* hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCP), linkage analysis, RNase protection assay, allele specific oligonucleotide (ASO), dot blot analysis and PCR-SSCP analysis. Also useful is the recently developed technique of DNA microchip technology. For example, either (1) the nucleotide sequence of both the cloned alleles and normal HPC2 gene or appropriate fragment (coding sequence or genomic sequence) are determined and then compared, or (2) the RNA transcripts of the HPC2 gene or gene fragment are hybridized to single stranded whole genomic DNA from an individual to be tested, and the resulting heteroduplex is treated with Ribonuclease A (RNase A) and run on a denaturing gel to detect the location of any mismatches. Two of these methods can be carried out according to the following procedures.

The alleles of the *HPC2* gene in an individual to be tested are cloned using conventional techniques. For example, a blood sample is obtained from the individual. The genomic DNA

isolated from the cells in this sample is partially digested to an average fragment size of approximately 20 kb. Fragments in the range from 18-21 kb are isolated. The resulting fragments are ligated into an appropriate vector. The sequences of the clones are then determined and compared to the normal *HPC2* gene.

Alternatively, polymerase chain reactions (PCRs) are performed with primer pairs for the 5' region or the exons of the *HPC2* gene. PCRs can also be performed with primer pairs based on any sequence of the normal *HPC2* gene. For example, primer pairs for one of the introns can be prepared and utilized. Finally, RT-PCR can also be performed on the mRNA. The amplified products are then analyzed by single stranded conformation polymorphisms (SSCP) using conventional techniques to identify any differences and these are then sequenced and compared to the normal gene sequence.

Individuals can be quickly screened for common *HPC2* gene variants by amplifying the individual's DNA using suitable primer pairs and analyzing the amplified product, e.g., by dot-blot hybridization using allele-specific oligonucleotide probes.

The second method employs RNase A to assist in the detection of differences between the normal *HPC2* gene and defective genes. This comparison is performed in steps using small (~500 bp) restriction fragments of the *HPC2* gene as the probe. First, the *HPC2* gene is digested with a restriction enzyme(s) that cuts the gene sequence into fragments of approximately 500 bp. These fragments are separated on an electrophoresis gel, purified from the gel and cloned individually, in both orientations, into an SP6 vector (e.g., pSP64 or pSP65). The SP6-based plasmids containing inserts of the *HPC2* gene fragments are transcribed *in vitro* using the SP6 transcription system, well known in the art, in the presence of [α -³²P]GTP, generating radiolabeled RNA transcripts of both strands of the gene.

Individually, these RNA transcripts are used to form heteroduplexes with the allelic DNA using conventional techniques. Mismatches that occur in the RNA:DNA heteroduplex, owing to sequence differences between the *HPC2* fragment and the *HPC2* allele subclone from the individual, result in cleavage in the RNA strand when treated with RNase A. Such mismatches can be the result of point mutations or small deletions in the individual's allele. Cleavage of the RNA strand yields two or more small RNA fragments, which run faster on the denaturing gel than the RNA probe itself.

Any differences which are found, will identify an individual as having a molecular variant of the *HPC2*. These variants can take a number of forms. The most severe forms would be frame shift mutations or large deletions which would cause the gene to code for an abnormal

protein or one which would significantly alter protein expression. Less severe disruptive mutations would include small in-frame deletions and nonconservative base pair substitutions which would have a significant effect on the protein produced, such as changes to or from a cysteine residue, from a basic to an acidic amino acid or vice versa, from a hydrophobic to hydrophilic amino acid or vice versa, or other mutations which would affect secondary or tertiary protein structure. Silent mutations or those resulting in conservative amino acid substitutions would not generally be expected to disrupt protein function.

Genetic testing will enable practitioners to identify individuals at risk prostate cancer, at, or even before, birth. Presymptomatic diagnosis of these epilepsies will enable prevention of these disorders.

EXAMPLES

The present invention is further detailed in the following Examples, which are offered by way of illustration and are not intended to limit the invention in any manner. Standard techniques well known in the art or the techniques specifically described below are utilized.

EXAMPLE 1

Genetic localization of HPC2

A set of high risk prostate cancer kindreds has been collected in Utah since 1990 for the purpose of localization of prostate cancer susceptibility loci. In Feb 1996, linkage analysis of data from a genome scan performed on a subset of the families noted evidence for linkage with markers on chromosome 17p. Subsequent analysis of more markers in this region of chromosome 17p in a larger set of families has led to strong linkage evidence for a susceptibility gene.

TABLE 1

Chromosome 17p two-point linkage evidence

Marker	17p map position	Heterogeneity Lod Score
D17S786	20.0	4.21
Myr 0022	25.5	3.99
Myr 0088	27.0	3.46
D17S947	31.6	2.32

Myr 0084	31.9	3.02
Myr 0079	32.0	0.99
D17S805	43.6	2.25

The study of specific kindreds with strong evidence of linkage to chromosome 17p allows the definition of a most likely region for the susceptibility locus by identifying the smallest inherited piece of chromosome 17p shared by the prostate cancer cases in the kindred.

5 The minimal genetically defined region is based on a telomeric recombinant in kindred 4325 and a centromeric recombinant in kindred 4320. Kindred 4325 was ascertained from a sibship of early onset prostate cancer cases. There are 6 affected brothers in this family, one of whom also has an affected son. Five of the 6 affected brothers, and the affected son, all share the same piece of chromosome 17p from somewhere below marker myr0065 down to and including
10 marker D17S805. Kindred 4320 was also ascertained from a sibship of early onset prostate cancer cases. In this kindred 3 affected brothers and an affected nephew share a piece of chromosome 17p from D17S786 down to and including myr0084. Together, the kindred 4325 and kindred 4320 recombinants define a minimal region of about 1 megabase (Figure 2A); this localization is well supported by a larger set of recombinants in both directions.

15 EXAMPLE 2

Contig assembly and genomic sequencing in the minimal genetically defined HPC2 region

Contig assembly. Given a genetically defined interval flanked by meiotic recombinants, one needs to generate a contig of genomic clones that spans that interval. Publicly available
20 resources, such as the Whitehead integrated maps of the human genome (e.g., the WICGR Chr 17 map) provide aligned chromosome maps of genetic markers, other sequence tagged sites (STSs), radiation hybrid map data, and CEPH yeast artificial chromosome (YAC) clones.

Oligonucleotide primer pairs for the markers located in the interval were synthesized and used to screen libraries of bacterial artificial chromosomes (BACs) to identify BACs in the
25 region. The initial set of markers used was D17S969, WI-2437, WI-2335, D17S947, and D17S799 (Figure 2A). BACs identified with these markers were end-sequenced. PCR primers designed from those end sequences were used as markers to arrange the initial BACs into contigs. The outermost marker from each contig was used in successive rounds of BAC library screening, eventually enabling the completion of a BAC clone contig that spanned the

genetically defined interval. A set of overlapping but non-redundant BAC clones that spanned this interval (Figure 2A) was then selected for use in subsequent molecular cloning protocols such as genomic sequencing.

Genomic sequencing. Given a tiling path of BAC clones across a defined interval, one useful gene finding strategy is to generate an almost complete genomic sequence of that interval. Two types of random genomic clone sublibraries were prepared from each BAC on the tiling path; these were Sau 3A partial digest libraries with inserts in the 5 to 8 kb size range, and random shear libraries with inserts in the 1.0 to 1.5 kb size range. Plasmid DNA from individual clones from the Sau 3A sublibraries sufficient in number to generate an, on average, 1x redundant sequence of each BAC was prepared using an Autogen robotic plasmid preparation machine (Integrated Separation Systems). Insert DNA from individual clones from the random shear sublibraries sufficient in number to generate an, on average, 5x redundant sequence of each BAC, was prepared by PCR with vector primers directly from aliquots of bacterial cultures of each individual clone. The resulting DNA templates were subjected to DNA sequencing from both ends with M13 forward or reverse fluorescent dye-labeled primers on ABI 377 sequencers.

These sequences were assembled into sequence contigs using the program Acem.bly (Thierry-Mieg et al., 1995; Durbin and Thierry-Mieg, 1991). The genomic sequence contigs were placed in a Genetic Data Environment (GDE) (Smith et al., 1994) local database for subsequent similarity searches. Similarities among genomic DNA sequences and GenBank entries - both DNA and protein - were identified using BLAST (Altschul et al., 1990). The DNA sequences were also characterized with respect to short period repeats, CpG content, and long open reading frames.

EXAMPLE 3

Sequence assembly of the human HPC2 gene

A BLAST search of genomic sequences from BAC 31k12 against dbEST identified two independent sets of human ESTs that, when parsed across the BAC 31k12 genomic sequences, revealed the presence of two independent multi-exon candidate genes, 04CG09 and the HPC2 gene (Figure 2A). A subset of the EST sequences assigned to HPC2 (Table 2) was assembled to produce a tentative partial cDNA sequence for the gene.

TABLE 2

Human ESTs used to assemble a tentative partial human HPC2 cDNA sequence

EST Accession #	Exon Span
AA679618	1→6
Z17886	4→8
W37591	7→12
AA310236	12→16
R55841	15→19
T34216	18→21
AA634909	20→24
AA504412	23→24
R42795	24→polyA

The individual exons of the human HPC2 gene were identified by parsing that tentative cDNA sequence across the BAC 31k12 genomic sequence (see schematics in Figures 2D and 2E). After we had identified the HPC2 gene, the MIT genome sequencing completely sequenced another BAC, 597m12, that also contains all of the exons of HPC2 (Genbank accession # AC005277). The sequence of the human HPC2 gene was corrected both by comparison of the sequences of the individual exons from the tentative cDNA assembly to the corresponding genomic sequences of BACs 31k12 and 597m12, and by mutation screening the gene from a set of human genomic DNAs (see Example 5).

The original tentative human HPC2 cDNA sequence contained neither the start codon nor any of the 5' UTR. These were obtained by 5' RACE. Briefly, a biotinylated reverse primer, CA4cg07.BR2, was designed from the sequence of the third exon of the human HPC2 gene and used, along with the anchor primer 5ampA, for a first round of PCR amplification from human fetal liver cDNA that had been prepared such that the 5' ends of cDNA molecules are anchored with the sequence 5tag1. The resulting PCR products were captured on streptavidin paramagnetic particles (Dynal), washed, and used as template in a second round PCR amplification. A phosphorylated reverse primer, CA4cg07.PR2, was designed from the sequence of the second exon of the human HPC2 sequence and used, along with the nested phosphorylated anchor primer 5ampB, for the second round PCR amplification. The resulting 5' RACE products were gel purified and sequenced with the primer CA4cg07.PR2 using dye-terminator chemistry and ABI 377 sequencers. Analysis of the sequences of these 5' RACE

products yielded both the start codon and part of the 5' UTR including an in-frame stop codon (Figure 3). Sequences of the primers used for 5' RACE are given in Table 4.

A full length human HPC2 cDNA was amplified from human head and neck cDNA using the primers CA4cg7.ATG and CA4cg7.TGA. The cDNA was ligated into the vector pGEM-T Easy (Promega) and transformed into E. coli. The sequence of the cDNA clone was confirmed by dye terminator sequencing on ABI 377 sequencers. Sequences of primers used to amplify the cDNA construct and confirm the sequence of the cDNA clone are also given in Table 3.

Table 3

**Primers used in 5' RACE, cDNA cloning, and
sequence confirmation of a full-length human HPC2 cDNA**

<u>5'RACE PRIMERS</u>	<u>Sequence (SEQ ID NO:)</u>
5tag1	CAG GAA TTC AGC ACA TAC TCA TTG TTC Agn n (29)
5AmpA	CAG GAA TTC AGC ACA TAC TCA (30)
5AmpB	(P)TT CAG CAC ATA CTC ATT GTT CA (31)
CA4cg07.BR2	(B)TG AAC GCC TTC TCC ACA GT (32)
CA4cg07.PR2	(P)GT ACC CGC TGC CAC CAC (33)
<u>EXPRESSION CONSTRUCT PRIMERS</u>	
CA4cg7.ATG	GCT AGG ATC CGC CAC CAT GTG GGC GCT TTG CTC (34)
CA4cg7.TGA	GCT ACT CGA GTC ACT GGG CTC TGA CCT TC (35)
<u>SEQUENCING PRIMERS</u>	
M13F20	GTA AAA CGA CGG CCA GT (36)
M13R20	GGA AAC AGC TAT GAC CAT G (37)
CA4cg7F1	TGC GCA CGC GAG AGA AG (38)
CA4cg7R1	CGC TTC TCT CGC GTG CG (39)
CA4cg7F2	TCT AAT GTT GGG GGC TTA (40)
CA4cg7R2	TAA GCC CCC AAC ATT AGA (41)
CA4cg7F3	TGA AAA TGA GCC ACA CCT (42)
CA4cg7R3	AGG TGT GGC TCA TTT TCA (43)
CA4cg7F4	CAT TCA ACC CAT CTG TGA (44)
CA4cg7R4	TCA CAG ATG GGT TGA ATG (45)
CA4cg7F5	TGA ATG CCT CCT CAA GTA (46)
CA4cg7R5	TAC TTG AGG AGG CAT TCA (47)
CA4cg7F6	GCT ACT GGA CTG TGG TGA (48)
CA4cg7R6	TCA CCA CAG TCC AGT AGC (49)
CA4cg7F7	TGG AAG AGT TTC AGA CCT G (50)
CA4cg7R7	CAG GTC TGA AAC TCT TCC A (51)
CA4cg7F8	CGC AGG GAC GCA CCA TA (52)

CA4cg7R8	GGT TGA ACT CGG AGA AGA (53)
CA4cg7F9	CAA CTG GAA AAA TAC CTC G (54)
CA4cg7F10	GCA GAG TCC AGA AAG GC (55)
CA4cg7F11	AGA GGA AAC TTC TTG GTG C (56)
CA4cg7F12	ACC AAG GAA AGG CAG ATG (57)
CA4cg7F13	GTC AAC ATA AGC CCC GAC (58)
CA4cg7F14	GGC TGC TGT GTT TGT GTC (59)
CA4cg7R14	GAA GGC ATT TGG CAG GA (60)
CA4cg7F15	TAT GAT TCC TGC CAA ATG (61)
CA4cg7R15	TCC AGC CAG AGG TGT GC (62)
CA4cg7F16	TGC GAG GCT CTG GTC CG (63)
CA4cg7R16	GGG CAT TGT TGG AAA GTC (64)
CA4cg7F17	TGT TTG CTG GCG ACA TC (65)

nn – the last 2 nucleotides of the anchor sequence 5tag1 are specific for each cDNA prep.

(P) indicates phosphate at the 5' end of the oligo

(B) indicates biotin at the 5' end of the oligo

EXAMPLE 4

Sequence assembly of the mouse HPC2 gene

A BLAST search of the assembled HPC2 cDNA sequence against dbEST identified 5 mouse ESTs that derived from a very similar gene, the mouse ortholog of HPC2, Mm.HPC2; their accession numbers are listed in Table 4.

Table 4

Mouse ESTs used to assemble a tentative partial Mm.HPC2 cDNA sequence

EST Accession #	Exon Span
AA563096	1→5
AA518169	8→14
AI132016	16→17
AA184645	19→24
AA174437	24→24

The original partial Mm.HPC2 cDNA sequence contained the start codon but little of the 5' UTR. More extensive 5' UTR sequence was obtained by 5' RACE. Briefly, a biotinylated reverse primer, m04cg07BR1, was designed from the sequence of the fourth exon of the mouse HPC2 gene and used, along with the anchor primer 5ampA, for a first round of PCR amplification from mouse embryo cDNA that had been prepared such that the 5' ends of cDNA molecules are anchored with the sequence 5tag1. The resulting PCR products were captured on

streptavidin paramagnetic particles (Dyna), washed, and used as template in a second round PCR amplification. A phosphorylated reverse primer, m04cg07PR1, was designed from the sequence of the third exon of the mouse HPC2 sequence and used, along with the nested phosphorylated anchor primer 5ampB, for the second round PCR amplification. The resulting 5' RACE products were gel purified and sequenced with the primers m04cg07PR1 and m04cg07 exon2 rev using dye-terminator chemistry and ABI 377 sequencers. Analysis of the sequences of these 5' RACE products yielded both the start codon and part of the 5' UTR including an in-frame stop codon (Figure 3). Sequences of the primers used for 5' RACE are given in Table 5.

More extensive 5' UTR sequence, sequence that may be from the promoter, and the sequences of intron 1 and intron 2 of the mouse HPC2 gene were obtained by genomic sequencing. BAC 428n12 was obtained from a mouse genomic library by screening the library by PCR with a pair of primers (04CG7.m11f1 and 04CG7.m11r1, Table 5) derived from exon 11 of the mouse HPC2 cDNA sequence. A primer pair derived from the SP6 end sequence of BAC 428n12 (428n12.S6.F1 and 428n12.S6.R1, Table 5) was used to screen the mouse BAC library by PCR; several overlapping BACs, including BAC 199n11, were identified. BACs 428n12 and 199n11 were sequenced with a series of 13 sequencing primers (mcg7f1 to mcg7r7, Table 5) derived from mouse HPC2 cDNA dye-terminator chemistry and ABI 377 sequencers. A subset of these sequences were assembled into a genomic sequence contig extending from 280 bp upstream of the ATG start codon of exon 1 into exon 3.

A full length mouse HPC2 cDNA is amplified from mouse embryo, placenta, or fetal brain cDNA using the primers msCA4cg7.f out and msCA4cg7.r out. The cDNA is reamplified with the primers msCA4cg7.ATG and msCA4cg7.TGA. The resulting PCR products are gel purified, ligated into the vector pGEM-T Easy (Promega), and transformed into E. coli. The sequence of the cDNA clone are confirmed dye terminator sequencing on ABI 377 sequencers.

Sequences of primers in use to amplify the cDNA construct are also given in Table 5.

Table 5

Primers used in 5' RACE and cDNA cloning of a full-length mouse HPC2 cDNA

<u>5'RACE PRIMERS</u>	Sequence (SEQ ID NO:)
5tag1	CAG GAA TTC AGC ACA TAC TCA TTG TTC Agn n (66)
5ampA	CAG GAA TTC AGC ACA TAC TCA (67)
5 Amp B	(P)TT CAG CAC ATA CTC ATT GTT CA (68)
m04cg07BR1	(B)CA GAA CAC ATT TGG GAA GC (69)
m04cg07PR1	(P)GA TGT TGT CCA AGC GAG C (70)

BAC library screening primers

04CG7.m11f1	TGA CAC ACA GCA CCT GA (71)
04CG7.m11r1	GAA GAT GTC AGG GTG GA (72)
428n12.S6.F1	CAG GCA TAC CAC TAC AGA (73)
428n12.S6.R1	TAT CAA CTT CTA GGC AAG TG (74)

Genomic sequencing primers

mcg7f1	GCA CCA TGT CGC AGG GTT C (75)
mcg7r1	GAA CCC TGC GAC ATG GTG C (76)
mcg7f2	TCG CAG GGT TCG GCT CGT C (77)
mcg7r2	AAC CCT GCG ACA TGG TGC G (78)
mcg7f3	AAA GAC CCA CTG CGA CAC C (79)
mcg7r3	GCA GGT GTC GCA GTG GGT C (80)
mcg7f4	CCG AAC ACC GTG TAC CTG CA (81)
mcg7r4	CAG GTA CAC GGT GTT CGG G (82)
mcg7f5	GTC TTC TCG GAA TAC AAC AGG (83)
mcg7r5	CTG TTG TAT TCC GAG AAG AC (84)
mcg7f6	AAG GCG TCC AAC GAC TTA TG (85)
mcg7r6	AGT CGT TGG ACG CCT TCT CC (86)
mcg7r7	TCC GAG TCA GAA AGA TGT TG (87)

EXPRESSION CONSTRUCT PRIMERSPRIMARY PCR

msCA4cg7.f out	GCC TTG TCA GCC TGG TG (88)
msCA4cg7.r out	AGG AAG TGA GCA GAG CG (89)

SECONDARY PCR

msCA4cg7.ATG	GCT AAA GCT TGC CAC CAT GTG GGC GCT CCG CTC (90)
msCA4cg7.TGA	GCT ACT CGA GTC ACA CTC GCG CTC CTA (91)

SEQUENCING PRIMERS

m04cg07 exon2 rev	GCC TTC TCC GCA GTT A (92)
-------------------	----------------------------

nn – the last 2 nucleotides of the anchor sequence 5tag1 are specific for each cDNA prep.
(P) indicates phosphate at the 5' end of the oligo
(B) indicates biotin at the 5' end of the oligo

EXAMPLE 5

Mutation screening of the human HPC2 gene

Using genomic DNAs from prostate kindred members, prostate cancer affecteds, and tumor cell lines as templates, nested PCR amplifications were performed to generate PCR products to screen for mutations in the HPC2 gene. The primers listed in Table 6 were used to amplify segments of the HPC2 gene. Using the outer primer pair for each amplicon (1A-1P, i.e.,

forward A and reverse P of amplicon 1), 10-20 ng of genomic DNA were subjected to a 25 cycle primary amplification, after which the PCR products were diluted 45-fold and reamplified using nested M13-tailed primers (1B-1Q, 1C-1R i.e., nested forward B and nested reverse Q of amplicon 1 or nested forward C and nested reverse R of amplicon 1) for another 23 cycles. In general, samples were amplified with Taq Platinum (Life Technologies) DNA polymerase; cycling parameters included an initial denaturation step at 95°C for 3 min, followed by cycles of denaturation at 96°C (12 s), annealing at 55°C (15 s) and extension at 72°C (30-60 s). After the PCR reactions, excess primers and deoxynucleotide triphosphates were digested with exonuclease I (United States Biochemicals) and shrimp alkaline phosphatase (Amersham). PCR products were sequenced with M13 forward or reverse fluorescent (Big Dye, ABI) dye-labeled primers on ABI 377 sequencers. Chromatograms were analyzed for the presence of polymorphisms or sequence aberrations in either the Macintosh program Sequencher (Gene Codes) or the Java program Mutscreen.

Table 6

Primers used to mutation screen the HPC2 gene from genomic DNA

Exon/Primer name	Sequence (SEQ ID NO:)
HPC2 exon 1	
ca4cg7.m1Anew	CCG CTT GAG ACG CTC TAG TAT (93)
ca4cg7.m1P	GCT CCG AAA GTG CTG ACA G (94)
ca4cg7.m1Bnew	GTT TTC CCA GTC ACG ACG TTT CTA TTG GAT GAG CAG CCT (95)
ca4cg7.m1Qnew	AGG AAA CAG CTA TGA CCA TGC CTG CGA TAT GGT GCG TC (96)
ca4cg7.m1C	GTT TTC CCA GTC ACG ACG CTC AGT TTT GGT GGA GAC G (97)
ca4cg7.m1Rnew	AGG AAA CAG CTA TGA CCA TGT GCC CCG ATG CTC AGA G (98)
HPC2 exons 2&3	
	(primary)
ca4cg7.m2&23 A2	AAT GGT GTC AGA GAG TTT ACA G (99)
ca4cg7.m2&23P	GCT ATT TGG GAG GCT GAG G (100)
HPC2 exon 2	
	(nested)
ca4cg7.m2B	GTT TTC CCA GTC ACG ACG AAT GGT GTC AGA GAG TTT ACA G (101)
ca4cg7.m2Q	AGG AAA CAG CTA TGA CCA TGA ACA AGG ACC ACT TTT GCT AT (102)
HPC2 exon 3	
	(nested)
ca4cg7.m23B	GTT TTC CCA GTC ACG ACG TTT ATA GCA AAA GTG GTC CTT G (103)
ca4cg7.m23Q	AGG AAA CAG CTA TGA CCA TGA GAC TTC CCA CCA GCC TC (104)
HPC2 exon 4	

ca4.cg07.m24A	CCT TGC TGC TTC ACC CTA G (105)
ca4.cg07.m24P	TGC TTT ATA TGT GCT GCT ACG (106)
ca4.cg07.m24B	GTT TTC CCA GTC ACG ACG CAT CTT CCC TGG TTG TAC TTC (107)
ca4.cg07.m24Q	AGG AAA CAG CTA TGA CCA TCT GGA GGG CAG AAG ACT GAT (108)
HPC2 exon 5	
ca4cg7.m3A	CTA CAT TTG TTC AAC CAT AAC TG (109)
ca4cg7.m3P	GAT TTT GAG GTT TGA TGT TGA TG (110)
ca4cg7.m3B	GTT TTC CCA GTC ACG ACG CAT TTG TTC AAC CAT AAC TGC (111)
ca4cg7.m3Q	AGG AAA CAG CTA TGA CCA TAT TTG AGA GGT CAG GGC ATA (112)
HPC2 exon 6	
ca4cg7.m4A	TCG TGT CAG ATT CCC ACC ATA (113)
ca4cg7.m4P	AGG CAT AAG TCA GAC ATC CGT (114)
ca4cg7.m4B	GTT TTC CCA GTC ACG ACG GTT ACT CTT CCC ACA CAT CTT C (115)
ca4cg7.m4Q	AGG AAA CAG CTA TGA CCA TCA CAG CAA GTG TTC AGT TTC TA (116)
HPC2 exon 7	
ca4cg7.m5A	CAT TCC CAT GTA TGA ACG TCT (117)
ca4cg7.m5P	ATA GTA AGC CCA GGA AGA AGGA (118)
ca4cg7.m5B	GTT TTC CCA GTC ACG ACG CAT TCC CAT GTA TGA ACG TCT (119)
ca4cg7.m5Q	AGG AAA CAG CTA TGA CCA TCT ACA AGC ATT ACA AGG CAG AG (120)
HPC2 exon 8	
ca4cg7.m6A	AGT GTC TTC AGC CTT TGT ATT G (121)
ca4cg7.m6P	ATC TGC TAT CTC TTC TTG TCT CA (122)
ca4cg7.m6B	GTT TTC CCA GTC ACG ACG ATC GGG TCA TAA TCA GTC TGT G (123)
ca4cg7.m6Q	AGG AAA CAG CTA TGA CCA TAT CTC TTC TTG TCT CAG GTA ACA (124)
HPC2 exons 9&10	
ca4cg7.m7&8A	(primary) CTT CTG AAA GCA ATA AAC GCA T (125)
ca4cg7.m7&8P	GAT GTC CAA ACT GTT CCA CG (126)
HPC2 exon 9	
ca4cg7.m7B	(nested) GTT TTC CCA GTC ACG ACG TAA AAC CAA CCT TCT TCA TTA G (127)
ca4cg7.m7Q	AGG AAA CAG CTA TGA CCA TAG CAA TGA TGG GAG CGA TG (128)
HPC2 exon 10	
ca4cg7.m8B	(nested) GTT TTC CCA GTC ACG ACG GGC TTC TGG GGA CTC ACT G (129)
ca4cg7.m8Q	AGG AAA CAG CTA TGA CCA TCC TTC AAA AGT GGT GTC TGT AG (130)
HPC2 exon 11	
ca4.cg07.m9A	GTA TCC ACA AAG AGA CCA GAA G (131)
ca4.cg07.m9P	CAC CAA CTA CCA ACA GTG ACT TA (132)
ca4.cg07.m9B	GTT TTC CCA GTC ACG ACG GCT CAC TGG ATA GGA TAT GTC AT (133)

ca4.cg07.m9Q	AGG AAA CAG CTA TGA CCA TCC AGA AAC ACA GCT CTT GCC (134)
HPC2 exon 12	
ca4.cg07.m10A	GCT TGC CAG ATA CAG GAA TC (135)
ca4.cg07.m10P	ACA GAA AGT TTA GGC AGG TG (136)
ca4.cg07.m10B	GTT TTC CCA GTC ACG ACG ACG ATA CCC CTC CCT GGC T (137)
ca4.cg07.m10Q	AGG AAA CAG CTA TGA CCA TAC AGA AAG TTT AGG CAG GTG (138)
HPC2 exons 13&14	(primary)
ca4.cg07.m11&12A	CCT CTC ACT CTT CCC AGC AC (139)
ca4.cg07.m11&12P	GGA GTA GGC TGC TTT TCT AAA T (140)
HPC2 exon 13	(nested)
ca4.cg07.m11B	GTT TTC CCA GTC ACG ACG GAA CAC CTC ATC CTC ATT ACC A (141)
ca4.cg07.m11Q	AGG AAA CAG CTA TGA CCA TAA GAG ACA AAA CAC ATT CAT GG (142)
HPC2 exon 14	(nested)
ca4.cg07.m12B	GTT TTC CCA GTC ACG ACG GTT TCC GCT GTA AGG TAG TGT (143)
ca4.cg07.m12Q	AGG AAA CAG CTA TGA CCA TCT GGA ACA TTT ACT ATG TGG CTA (144)
HPC2 exon 15	
ca4.cg07.m13A	TGC TAG TGG GTA GAG GTC AG (145)
ca4.cg07.m13P	ACT GAA AGC CAG GTT AGA ATG (146)
ca4.cg07.m13B	GTT TTC CCA GTC ACG ACG ACC CTG TCC GTC ACC TGA G (147)
ca4.cg07.m13Q	AGG AAA CAG CTA TGA CCA TCC CAC CAG CAC TCC ACT TA (148)
HPC2 exon 16	
ca4cg07.m14A	TGT GAA GAC GGG ATA ACC TGA (149)
ca4cg07.m14P	GAC AGG GCT TGA TAC CGCA (150)
ca4cg07.m14B	GTT TTC CCA GTC ACG ACG ATG CTG GCT CAC TTT TGA CC (151)
ca4cg07.m14Q	AGG AAA CAG CTA TGA CCA TGAC TGG TGA GTA CAG CAG GA (152)
HPC2 exon 17	
ca4.cg07.m15A	CCA GCC TTT GTG TAA GTC TAC (153)
ca4.cg07.m15P	TCT GGG CAA GTT TGG AAG C (154)
ca4.cg07.m15B	GTT TTC CCA GTC ACG ACG TCC AAA GCA GAC ATC AGC CTC (155)
ca4.cg07.m15Q	AGG AAA CAG CTA TGA CCA TGG AGG AAA AGA CGC AGC CA (156)
HPC2 exon 18	
ca4.cg07.m16A	CGC TTT CTG CCT GTG ACA T (157)
ca4.cg07.m16P	TTC TGT CCT TCA GCC AAT GC (158)
ca4.cg07.m16B	GTT TTC CCA GTC ACG ACG TTA GAG GCT GGT GGG TGA C (159)
ca4.cg07.m16Q	AGG AAA CAG CTA TGA CCA TCA TCT CAA TAA AAA CTG GAG TGC (160)
HPC2 exon 19	

ca4.cg07.m17A	CAC TTG ATG GGC GTT CTG AG (161)
ca4.cg07.m17P	TTC TGT CCT TCA GCC AAT GC (162)
ca4.cg07.m17B	GTT TTC CCA GTC ACG ACG TTC CAG CGG TTT ACA CAT CA (163)
ca4.cg07.m17Q	AGG AAA CAG CTA TGA CCA TTA CCC CAG TGT CCA CCT TG (164)
HPC2 exons 20&21	(primary)
CA4CG7.m18&22A	GGG TTC TCC AGC CAA AGA CT (165)
CA4CG7.m18&22P	CTG AGT CTC CTG CCT CTG C (166)
HPC2 exon 20	(nested)
ca4.cg07.m18B	GTT TTC CCA GTC ACG ACG GGG TTC TCC AGC CAA AGA CT (167)
ca4.cg07.m18Q	AGG AAA CAG CTA TGA CCA TGT GGG GCT GGA AGG CTC TG (168)
HPC2 exon 21	(nested)
ca4.cg07.m22B	GTT TTC CCA GTC ACG ACG AAG AGG TAA GGG GCA CAG C (169)
ca4.cg07.m22Q	AGG AAA CAG CTA TGA CCA TCT GAG TCT CCT GCC TCT GC (170)
HPC2 exon 22	
ca4.cg07.m19A	GCT GAG TGT TGA GAC CAG GA (171)
ca4.cg07.m19P	AGA CAA ACG ACG GCT GCT C (172)
ca4.cg07.m19B	GTT TTC CCA GTC ACG ACG TTG AGA CCA GGA AAC AGC AC (173)
ca4.cg07.m19Q	AGG AAA CAG CTA TGA CCA TGA GAG GAT GTG GGC GAC AA (174)
HPC2 exon 23	
ca4.cg07.m20A	GGG AGA TGG TGC TGG CTA C (175)
ca4.cg07.m20P	CCT GGT TAG TGA TGG GTA GAT (176)
ca4.cg07.m20B	GTT TTC CCA GTC ACG ACG CAG GGT CTG TGC CAC TGT C (177)
ca4.cg07.m20Q	AGG AAA CAG CTA TGA CCA TCT CAG TGT GTA GAG TCC TGT C (178)
HPC2 exon 24	splice acceptor and open reading frame
ca4.cg07.m21A	TTG ATT TTG AGA GCA TCT GGA C (179)
ca4.cg07.m21P	CTC GGA CAC TTA GAC CCA CTG (180)
ca4.cg07.m21B1	GTT TTC CCA GTC ACG ACG TGC ATC CCT TCC AGC TCC T (181)
ca4.cg07.m21Q	AGG AAA CAG CTA TGA CCA TGA CAC ACA GCC TTC TGA GTT CA (182)
ca4.cg07.m21C	GTT TTC CCA GTC ACG ACG CCA CAC AGA GGA GCC ACA G (183)
ca4.cg07.m21R	AGG AAA CAG CTA TGA CCA TAC CAG TCC TAA GAG GCA TCT ATA (184)
HPC2 exon 24	3' untranslated region
ca4.cg07.m21.3'UTR A	CCA CAC AGA GGA GCC ACA G (185)
ca4.cg07.m21.3'UTR P	CCA GAG GTG CTC ACT ACG AC (186)
ca4.cg07.m21.3'UTR B	GTT TTC CCA GTC ACG ACG AGG TCA GAG CCC AGT GAA GAT (187)
ca4.cg07.m21.3'UTR Q	AGG AAA CAG CTA TGA CCA TCA TCT GCT TGC TTC CGT GTG (188)
ca4.cg07.m21.3'UTR C	GTT TTC CCA GTC ACG ACG TCA GGA TAG GTG GTA TGG AGC (189)
ca4.cg07.m21.3'UTR R	AGG AAA CAG CTA TGA CCA TCG GAC ACT TAG ACC CAC TGA T (190)

Table 7
Sequence Variants

Variant name	Sequence (SEQ ID NO:)	Coding effect*
C650T	AGACTCCGAGTYGAATGAAATG (191)	Ser217Leu
A1580G	GGTGAGGGCACRITTTGGGCAGCT (192)	Thr520Thr
G1621A	GCACCTGGCTRCTGTGTTGTG (193)	Ala541Thr
1641insG (normal)	GTGTCCACCTG-CACGCAGATCA (194)	
(with insertion of G)	GTGTCCACCTGGCACGCAGATCA (195)	frameshift
C1722T	AAGCCGCTTCAYCCTTTGCTGGT (196)	His574His
A1893G	GCTGTTGCGAACRTGTGATTGGA (197)	Thr631Thr
C2632G	GAGGCTTGGSTCCACATAAG (198)	
C2687T	CCTGGCACAGCYGGGCCAGGA (199)	
G2801A	AATCCAGCAAARTGATTCCTGC (200)	
IVS2 T-11C	Taaatgttttcatctttag (201)	
IVS5 T-14C	Ttgctgttglyggtttctgt (202)	
IVS10 23insGAT (normal)	ggttttctgtat---tcagcagttaca (203)	
(with insertion of GAT)	Ggttttctgtatgattcagcagttaca (204)	
IVS13 C15T	Gtgtctcagacyggcccttgtc (205)	
IVS14 A17T	Tgccatcttgawctaataatggaac (206)	
IVS14 T-8C	Cttctctctctyctgcaggat (207)	
IVS16 C41T	Catcaagggcaygtttactttt (208)	
IVS19 C26G	Cagccitgcccctgggctgtg (209)	

*based on conceptual translation of the HPC2 ORF for each allele of the sequence variant.

In the course of mutation screening HPC2 from lymphocyte DNA collected from prostate cancer affecteds in our set of linked families, a frameshift mutation (1641insG) was detected in two affected brothers in kindred 4102. Individual 4102.001, diagnosed with prostate cancer at age 51, and his brother 4102.013, diagnosed with prostate cancer at age 46, both carry the frameshift. Further mutation screening in this kindred revealed that their mother, 4102.002, carries the frameshift. In addition, her maternal uncle 4102.053, diagnosed with prostate cancer at age 75 and deceased at age 76, also carried the frameshift.

The frameshift 1641insG throws translation of HPC2 out of frame following amino acid leucine547. Because the HPC2 protein shares significant sequence similarity with homologous proteins as evolutionarily distant as the E. coli *elaC* protein, and this region of shared sequence similarity extends over a segment of more than 180 amino acids downstream of leucine547,

there is no doubt but that the frameshift 1641insG is deleterious to the function of the HPC2 protein.

Taken together, the observation that the frameshift HPC2 1641insG segregates with prostate cancer across three generations of kindred 4102, and the inference from shared
 5 sequence similarity that the frameshift HPC2 1641insG must be deleterious to the function of the HPC2 protein, establish that deleterious germline mutations in the HPC2 gene confer susceptibility to prostate cancer.

EXAMPLE 6

Identification of HPC2-interacting proteins by two-hybrid analysis

DNA fragments encoding all or portions of HPC2 are ligated to a two-hybrid DNA-binding domain vector such as pGBT.C such that the coding sequence of *HPC2* is in-frame with coding sequence for the Gal4p DNA-binding domain. A plasmid that encodes a DNA-binding domain fusion to a fragment of HPC2 is introduced into the yeast reporter strain (such as J692) along with a
 15 library of cDNAs fused to an activation domain. Transformants are spread onto 20 - 150 mm plates of selective media, such as yeast minimal media lacking leucine, tryptophan, and histidine, and containing 25 mM 3-amino-1,2,4-triazole. After one week incubation at 30° C, yeast colonies are assayed for expression of the *lacZ* reporter gene by beta-galactosidase filter assay. Colonies that both grow in the absence of histidine and are positive for production of beta-galactosidase are
 20 chosen for further characterization.

The activation domain plasmid is purified from positive colonies by the smash-and-grab technique. These plasmids are introduced into *E. coli* (e.g., DH10B (Gibco BRL) by electroporation and purified from *E. coli* by the alkaline lysis method. To test for the specificity of the interaction, specific activation domain plasmids are cotransformed into strain J692 with
 25 plasmids encoding various DNA-binding domain fusion proteins, including fusions to segments of HPC2 and human lamin C. Transformants from these experiments are assayed for expression of the *HIS3* and *lacZ* reporter genes. Positives that express reporter genes with *Hs.HPC2* constructs and not with lamin C constructs encode bona fide HPC2-interacting proteins. These proteins are identified and characterized by sequence analysis of the insert of the appropriate activation domain
 30 plasmid.

EXAMPLE 7

Identification of orthologs of the human HPC2 gene

All species living on the Earth now are thought to have evolved from a single common ancestor that lived in the distant past, perhaps 3.5 to 4 billion years ago. This means that any pair of species must share a common ancestor species that lived at some time in the past. Admittedly, this view is a bit simplistic because, for instance, the nuclear genomes and mitochondrial genomes of eukaryotes are thought to have independent prokaryotic ancestries. During the evolution of an ancestral species into two or more extant daughter species, the genes present in the genome of the ancestral species evolve into the genes present in the genomes of the daughter species. The evolutionary history of the genes present in the daughter species can be quite complex because the individual genes can evolve through a diverse set of processes including nucleotide substitution, insertion, deletion, gene duplication, gene conversion, lateral transfer, etc. Even so, the evolutionary history of related genes in related organisms can often be sorted out, especially if the pair/set of species share a relatively recent common ancestor or if the genes being analyzed evolved primarily through nucleotide substitutions and/or small insertions and/or small deletions, but not gene duplications or gene conversions. When, upon analysis, it appears that a single gene in one species and a single gene in another species have evolved from a single gene in a common ancestor species, those genes are termed orthologs.

Knowledge of the identity of genes orthologous to disease-related human genes can often be quite useful.

EXAMPLE 8

Analysis of the HPC2 Gene

The structure and function of HPC2 gene are determined according to the following methods.

Biological Studies. Mammalian expression vectors containing HPC2 cDNA are constructed and transfected into appropriate prostate carcinoma cells with lesions in the gene. Wild-type HPC2 cDNA as well as altered HPC2 cDNA are utilized. The altered HPC2 cDNA can be obtained from altered HPC2 alleles or produced as described below. Phenotypic reversion in cultures (e.g., cell morphology, doubling time, anchorage-independent growth) and in animals (e.g., tumorigenicity) is examined. The studies will employ both wild-type and mutant forms of the gene.

Molecular Genetics Studies. *In vitro* mutagenesis is performed to construct deletion mutants and missense mutants (by single base-pair substitutions in individual codons and alanine scanning mutagenesis). The mutants are used in biological, biochemical and biophysical studies.

5 Mechanism Studies. The ability of HPC2 protein to bind to known and unknown DNA sequences is examined. Its ability to transactivate promoters is analyzed by transient reporter expression systems in mammalian cells. Conventional procedures such as particle-capture and yeast two-hybrid system are used to discover and identify any functional partners. The nature and functions of the partners are characterized. These partners in turn are targets for drug
10 discovery.

Structural Studies. Recombinant proteins are produced in *E. coli*, yeast, insect and/or mammalian cells and are used in crystallographical and NMR studies. Molecular modeling of the proteins is also employed. These studies facilitate structure-driven drug design.

15 EXAMPLE 9

Generation of Polyclonal Antibody against HPC2

Segments of HPC2 coding sequence are expressed as fusion protein in *E. coli*. The overexpressed proteins are purified by gel elution and used to immunize rabbits and mice using a procedure similar to the one described by Harlow and Lane, 1988. This procedure has been
20 shown to generate Abs against various other proteins (for example, see Kraemer, *et al.*, 1993).

Briefly, a stretch of HPC2 coding sequence was cloned as a fusion protein in plasmid PET5A (Novagen, Inc., Madison, WI). The HPC2 incorporated sequences might include SEQ ID NOs:1, 3 or 28 or portions thereof. After induction with IPTG, the overexpression of a fusion protein with the expected molecular weight is verified by SDS/PAGE. Fusion proteins
25 are purified from the gel by electroelution. The identification of the protein as the HPC2 fusion product is verified by protein sequencing at the N-terminus. Next, the purified protein is used as immunogen in rabbits. Rabbits are immunized with 100 µg of the protein in complete Freund's adjuvant and boosted twice in 3 week intervals, first with 100 µg of immunogen in incomplete Freund's adjuvant followed by 100 µg of immunogen in PBS. Antibody containing serum is
30 collected two weeks thereafter.

This procedure can be repeated to generate antibodies against mutant forms of the HPC2 protein. These antibodies, in conjunction with antibodies to wild type HPC2, are used to detect the presence and the relative level of the mutant forms in various tissues and biological fluids.

EXAMPLE 10

Generation of Monoclonal Antibodies Specific for HPC2

Monoclonal antibodies are generated according to the following protocol. Mice are immunized with immunogen comprising intact HPC2 or HPC2 peptides (wild type or mutant) conjugated to keyhole limpet hemocyanin using glutaraldehyde or EDC as is well known.

The immunogen is mixed with an adjuvant. Each mouse receives four injections of 10 to 100 µg of immunogen and after the fourth injection blood samples are taken from the mice to determine if the serum contains antibody to the immunogen. Serum titer is determined by ELISA or RIA. Mice with sera indicating the presence of antibody to the immunogen are selected for hybridoma production.

Spleens are removed from immune mice and a single cell suspension is prepared (see Harlow and Lane, 1988). Cell fusions are performed essentially as described by Kohler and Milstein, 1975. Briefly, P3.65.3 myeloma cells (American Type Culture Collection, Rockville, MD) are fused with immune spleen cells using polyethylene glycol as described by Harlow and Lane, 1988. Cells are plated at a density of 2×10^5 cells/well in 96 well tissue culture plates. Individual wells are examined for growth and the supernatants of wells with growth are tested for the presence of HPC2 specific antibodies by ELISA or RIA using wild type or mutant HPC2 target protein. Cells in positive wells are expanded and subcloned to establish and confirm monoclonality.

Clones with the desired specificities are expanded and grown as ascites in mice or in a hollow fiber system to produce sufficient quantities of antibody for characterization and assay development.

EXAMPLE 11

Isolation of HPC2 Binding Peptides

Peptides that bind to the HPC2 gene product are isolated from both chemical and phage-displayed random peptide libraries as follows.

Fragments of the HPC2 gene product are expressed as GST and His-tag fusion proteins in both *E. coli* and SF9 cells. The fusion protein is isolated using either a glutathione matrix (for GST fusions proteins) or nickel chelation matrix (for His-tag fusion proteins). This target fusion protein preparation is either screened directly as described below, or eluted with glutathione or imidazole. The target protein is immobilized to either a surface such as polystyrene; or a resin

such as agarose; or solid supports using either direct absorption, covalent linkage reagents such as glutaraldehyde, or linkage agents such as biotin-avidin.

Two types of random peptide libraries of varying lengths are generated: synthetic peptide libraries that may contain derivatized residues, for example by phosphorylation or myristylation, and phage-displayed peptide libraries which may be phosphorylated. These libraries are incubated with immobilized HPC1 gene product in a variety of physiological buffers. Next, unbound peptides are removed by repeated washes, and bound peptides recovered by a variety of elution reagents such as low or high pH, strong denaturants, glutathione, or imidazole. Recovered synthetic peptide mixtures are sent to commercial services for peptide micro-sequencing to identify enriched residues. Recovered phage are amplified, rescreened, plaque purified, and then sequenced to determine the identity of the displayed peptides.

Use of HPC1 binding peptides. Peptides identified from the above screens are synthesized in larger quantities as biotin conjugates by commercial services. These peptides are used in both solid and solution phase competition assays with HPC1 and its interacting partners identified in yeast 2-hybrid screens. Versions of these peptides that are fused to membrane-permeable motifs (Lin et al., 1995; Rojas et al., 1996) will be chemically synthesized, added to cultured cells and the effects on growth, apoptosis, differentiation, cofactor response, and internal changes will be assayed.

EXAMPLE 12

Sandwich Assay for HPC2

Monoclonal antibody is attached to a solid surface such as a plate, tube, bead, or particle. Preferably, the antibody is attached to the well surface of a 96-well ELISA plate. 100 μ l sample (e.g., serum, urine, tissue cytosol) containing the HPC2 peptide/protein (wild-type or mutant) is added to the solid phase antibody. The sample is incubated for 2 hrs at room temperature. Next the sample fluid is decanted, and the solid phase is washed with buffer to remove unbound material. 100 μ L of a second monoclonal antibody (to a different determinant on the HPC2 peptide/protein) is added to the solid phase. This antibody is labeled with a detector molecule (e.g., 125-I, enzyme, fluorophore, or a chromophore) and the solid phase with the second antibody is incubated for two hrs at room temperature. The second antibody is decanted and the solid phase is washed with buffer to remove unbound material.

The amount of bound label, which is proportional to the amount of HPC2 peptide/protein present in the sample, is quantitated. Separate assays are performed using monoclonal

antibodies which are specific for the wild-type HPC2 as well as monoclonal antibodies specific for each of the mutations identified in HPC2.

EXAMPLE 13

Two-hybrid Assay to Identify Proteins that Interact with HPC2

Sequence encoding all or portions of HPC2 are ligated to pAS2-1 (Clontech) such that the coding sequence of HPC2 is in-frame with coding sequence for the GAL4p DNA-binding domain. This plasmid construct is introduced into the yeast reporter strain Y190 by transformation. A library of activation domain fusion plasmids prepared from human prostate cDNA (Clontech) is then introduced into strain Y190 carrying the pAS2-1-based fusion construct. Transformants are spread onto 20 - 150 mm plates of yeast minimal media lacking leucine, tryptophan, and histidine, and containing 25 mM 3-amino-1,2,4-triazole. After one week incubation at 30°C, yeast colonies are assayed for expression of the lacZ reporter gene by β-galactosidase filter assay. Colonies that both grow in the absence of histidine and are positive for production of β-galactosidase are chosen for further characterization.

The activation domain plasmid is purified from positive colonies by the smash-and-grab technique. These plasmids are introduced into *E. coli* DH5α by electroporation and purified from *E. coli* by the alkaline lysis method. To test for the specificity of the interaction, specific activation domain plasmids are cotransformed into strain Y190 with plasmids encoding various DNA-binding domain fusion proteins, including fusions to HPC2 and human lamin C. Transformants from these experiments are assayed for expression of the HIS3 and lacZ reporter genes. Positives that express reporter genes with HPC2 constructs and not with lamin C constructs encode bona fide HPC2 interacting proteins. These proteins are identified and characterized by sequence analysis of the insert of the appropriate activation domain plasmid.

This procedure is repeated with mutant forms of the HPC2 gene, to identify proteins that interact with only the mutant protein or to determine whether a mutant form of the HPC2 protein can or cannot interact with a protein known to interact with wild-type HPC2.

While the invention has been disclosed in this patent application by reference to the details of preferred embodiments of the invention, it is to be understood that the disclosure is intended in an illustrative rather than in a limiting sense, as it is contemplated that modifications will readily occur to those skilled in the art, within the spirit of the invention and the scope of the appended claims.

LIST OF REFERENCES

- Altschul SF, *et al.* (1990). *J. Mol. Biol.* **215**: 403-410.
- Altschul SF, *et al.* (1997). *Nucl. Acids Res.* **25**:3389-3402.
- 5 Anand R (1992). Techniques for the Analysis of Complex Genomes. (Academic Press).
- Anderson WF, *et al.* (1980). *Proc. Natl. Acad. Sci. USA* **77**:5399-5403.
- Ausubel FM, *et al.* (1992). Current Protocols in Molecular Biology, (J. Wiley and Sons, NY).
- Bandyopadhyay PK and Temin HM (1984). *Mol. Cell. Biol.* **4**:749-754.
- Bartel PL, *et al.* (1993). "Using the 2-hybrid system to detect protein-protein interactions." *In*:
10 Cellular Interactions in Development: A Practical Approach, Oxford University Press,
pp. 153-179.
- Beaucage SL and Caruthers MH (1981). *Tetra. Letts.* **22**:1859-1862.
- Berglund P, *et al.* (1993). *Biotechnology* **11**:916-920.
- Berkner KL (1992). *Curr. Top. Microbiol. Immunol.* **158**:39-66.
- 15 Berkner KL, *et al.* (1988). *BioTechniques* **6**:616-629.
- Borman S (1996). *Chemical & Engineering News*, December 9 issue, pp. 42-43.
- Bouchardy C, *et al.* (1998). *Pharmacogenetics* **8**:291-298.
- Breakefield XO and Geller AI (1987). *Mol. Neurobiol.* **1**:337-371.
- Brinster RL, *et al.* (1981). *Cell* **27**:223-231.
- 20 Buchschacher GL and Panganiban AT (1992). *J. Virol.* **66**:2731-2739.
- Cannon L, *et al.* (1982). *Cancer Surveys* **1**:47-69.
- Capecci MR (1989). *Science* **244**:1288-1292.
- Cariello NF (1988). *Am. J. Human Genetics* **42**:726-734.
- Carter BS, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:3367-3371.
- 25 Carter BS, *et al.* (1993). *J. Urol.* **150**:797-802.
- Chee M, *et al.* (1996). *Science* **274**:610-614.
- Chevray PM and Nathans DN (1992). *Proc. Natl. Acad. Sci. USA* **89**:5789-5793.
- Compton J (1991). *Nature* **350**:91-92.
- Conner BJ, *et al.* (1983). *Proc. Natl. Acad. Sci. USA* **80**:278-282.
- 30 Costantini F and Lacy E (1981). *Nature* **294**:92-94.
- Cotton M, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:4033-4037.
- Cotton RG, *et al.* (1988). *Proc. Natl. Acad. Sci. USA* **85**:4397-4401.
- Culver KW, *et al.* (1992). *Science* **256**:1550-1552.
- Curiel DT, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:8850-8854.
- 35 Curiel DT, *et al.* (1992). *Hum. Gene Ther.* **3**:147-154.

- DeRisi J, et al. (1996). *Nature Genetics* **14**:457-460.
- Deutscher, M (1990). *Meth. Enzymology* **182**:83-89 (Academic Press, San Diego, Cal.).
- Donehower LA, et al. (1992). *Nature* **356**:215-221.
- 5 Durbin R and Thierry-Mieg J (1991). A C. *elegans* Database. Documentation, code and data available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and.ncbi.nlm.nih.gov.
- Editorial (1996). *Nature Genetics* **14**:367-370.
- Elghanian R, et al. (1997). *Science* **277**:1078-1081.
- 10 Enhancers and Eukaryotic Gene Expression. Cold Spring Harbor Press, Cold Spring Harbor, New York (1983).
- Erickson J, et al. (1990). *Science* **249**:527-533.
- Fahy E, et al. (1991). *PCR Methods Appl.* **1**:25-33.
- Feil et al., (1996). *Proc. Natl. Acad. Sci. USA* **93**:10887-10890.
- Felgner PL, et al. (1987). *Proc. Natl. Acad. Sci. USA* **84**:7413-7417.
- 15 Fields S and Song O-K (1989). *Nature* **340**:245-246.
- Fiers W, et al. (1978). *Nature* **273**:113-120.
- Fincham SM, et al. (1990). *The Prostate* **17**:189-206.
- Fink DJ, et al. (1992). *Hum. Gene Ther.* **3**:11-19.
- Fink DJ, et al. (1996). *Ann. Rev. Neurosci.* **19**:265-287.
- 20 Finkelstein J, et al. (1990). *Genomics* **7**:167-172.
- Fodor SPA (1997). *Science* **277**:393-395.
- Freese A, et al. (1990). *Biochem. Pharmacol.* **40**:2189-2199.
- Friedman T (1991). In: Therapy for Genetic Diseases, T. Friedman, ed., Oxford University Press, pp. 105-121.
- 25 Gagnet et al. (1997). *Nucl. Acids Res.* **25**:3326-3331.
- Giovannucci E, et al. (1997). *Proc. Natl. Acad. Sci. USA* **94**:3320-3323.
- Glover D (1985). DNA Cloning, I and II (Oxford Press).
- Goding (1986). Monoclonal Antibodies: Principles and Practice, 2d ed. (Academic Press, NY).
- Godowski PJ, et al. (1988). *Science* **241**:812-816.
- 30 Goldgar DE, et al. (1994). *J. Natl. Can. Inst.* **86**:3:200-209.
- Gordon JW, et al. (1980). *Proc. Natl. Acad. Sci. USA* **77**:7380-7384.
- Gordon JW (1989). *Intl. Rev. Cytol.* **115**:171-229.
- Gorziglia M and Kapikian AZ (1992). *J. Virol.* **66**:4407-4412.
- Graham FL and van der Eb AJ (1973). *Virology* **52**:456-467.
- 35 Grompe M (1993). *Nature Genetics* **5**:111-117.
- Grompe M, et al. (1989). *Proc. Natl. Acad. Sci. USA* **86**:5855-5892.

- Gu H, *et al.* (1994). *Science* **265**:103-106.
- Guthrie G and Fink GR (1991). Guide to Yeast Genetics and Molecular Biology (Academic Press).
- Hacia JG, *et al.* (1996). *Nature Genetics* **14**:441-447.
- 5 Harlow E and Lane D (1988). Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Harty LC, *et al.* (1997). *J. Natl. Cancer Inst.* **89**:1698-1705.
- Hasty P, *et al.* (1991). *Nature* **350**:243-246.
- Helseth E, *et al.* (1990). *J. Virol.* **64**:2416-2420.
- 10 Hodgson J (1991). *Bio/Technology* **9**:19-21.
- Hori H, *et al.* (1997). *J. Clin. Gastroenterol.* **25**:568-575.
- Huse WD, *et al.* (1989). *Science* **246**:1275-1281.
- Innis MA, *et al.* (1990). PCR Protocols: A Guide to Methods and Applications (Academic Press, San Diego, CA).
- 15 Jablonski E, *et al.* (1986). *Nucl. Acids Res.* **14**:6115-6128.
- Jakoby WB and Pastan IH (eds.) (1979). Cell Culture. Methods in Enzymology, Vol. 58 (Academic Press, Inc., Harcourt Brace Jovanovich (NY)).
- Johnson PA, *et al.* (1992). *J. Virol.* **66**:2952-2965.
- Johnson, *et al.* (1993). "Peptide Turn Mimetics" In: Biotechnology and Pharmacy, Pezzuto et al., eds., Chapman and Hall, NY.
- 20 Kaneda Y, *et al.* (1989). *J. Biol. Chem.* **264**:12126-12129.
- Kanehisa M (1984). *Nucl. Acids Res.* **12**:203-213.
- Kinszler KW, *et al.* (1991). *Science* **251**:1366-1370.
- Kohler G and Milstein C (1975). *Nature* **256**:495-497.
- 25 Krain LS (1974). *Preventive Medicine* **3**:154-159.
- Kubo T, *et al.* (1988). *FEBS Lett.* **241**:119-125.
- Kyte J and Doolittle RF (1982). *J. Mol. Biol.* **157**:105-132.
- Landegren U, *et al.* (1988). *Science* **242**:229-237.
- Lasko M, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:6232-6236.
- 30 Lavitrano M, *et al.* (1989). *Cell* **57**:717-723.
- Lee JE, *et al.* (1995). *Science* **268**:836-844.
- Lim CS, *et al.* (1991). *Circulation* **83**:2007-2011.
- Lin YZ, *et al.* (1995). *J. Biol. Chem.* **270**:14255-14258.
- Lipshutz RJ, *et al.* (1995). *BioTechniques* **19**:442-447.
- 35 Lo CW (1983). *Mol. Cell. Biol.* **3**:1803-1814.
- Lobe and Nagy (1998). *Bioessays* **20**:200-208.

- Lockhart DJ, et al. (1996). *Nature Biotechnology* **14**:1675-1680.
- Madzak C, et al. (1992). *J. Gen. Virol.* **73**:1533-1536.
- Maniatis T, et al. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 5 Mann R and Baltimore D (1985). *J. Virol.* **54**:401-407.
- Margolskee RF (1992). *Curr. Top. Microbiol. Immunol.* **158**:67-95.
- Martin R, et al. (1990). *BioTechniques* **9**:762-768.
- Matteucci MD and Caruthers MH (1981). *J. Am. Chem. Soc.* **103**:3185.
- Matthews JA and Kricka LJ (1988). *Anal. Biochem.* **169**:1.
- 10 Meikle AW, et al. (1985). *Prostate* **6**:121-128.
- Merrifield B (1963). *J. Am. Chem. Soc.* **85**:2149-2156.
- Metzger D, et al. (1988). *Nature* **334**:31-36.
- Mifflin TE (1989). *Clinical Chem.* **35**:1819-1825.
- Miller AD (1992). *Curr. Top. Microbiol. Immunol.* **158**:1-24.
- 15 Miller AD, et al. (1985). *Mol. Cell. Biol.* **5**:431-437.
- Miller AD, et al. (1988). *J. Virol.* **62**:4337-4345.
- Modrich P (1991). *Ann. Rev. Genet.* **25**:229-253.
- Mombaerts P, et al. (1992). *Cell* **68**:869-877.
- Morganti G, et al. (1956). *Acta Geneticae Medicae et Gemellogogiae* **6**:304-305.
- 20 Moss B (1992). *Curr. Top. Microbiol. Immunol.* **158**:25-38.
- Moss B (1996). *Proc. Natl. Acad. Sci. USA* **93**:11341-11348.
- Muzyczka N (1992). *Curr. Top. Microbiol. Immunol.* **158**:97-129.
- Nabel (1992). *Hum. Gene Ther.* **3**:399-410.
- Nabel EG, et al. (1990). *Science* **249**:1285-1288.
- 25 Naldini L, et al. (1996). *Science* **272**:263-267.
- Newton CR, et al. (1989). *Nucl. Acids Res.* **17**:2503-2516.
- Nguyen Q, et al. (1992). *BioTechniques* **13**:116-123.
- Novack DF, et al. (1986). *Proc. Natl. Acad. Sci. USA* **83**:586-590.
- Ohi S, et al. (1990). *Gene* **89**:279-282.
- 30 Orita M, et al. (1989). *Proc. Natl. Acad. Sci. USA* **86**:2776-2770.
- Osterrieder and Wolf (1998). *Rev. Sci. Tech.* **17**:351-364.
- Page KA, et al. (1990). *J. Virol.* **64**:5270-5276.
- Pellicer A, et al. (1980). *Science* **209**:1414-1422.
- Petropoulos CJ, et al. (1992). *J. Virol.* **66**:3391-3397.
- 35 Philpott KL, et al. (1992). *Science* **256**:1448-1452.

- Quantin B, *et al.* (1992). *Proc. Natl. Acad. Sci. USA* **89**:2581-2584.
- Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA).
- Rigby PWJ, *et al.* (1977). *J. Mol. Biol.* **113**:237-251.
- Rojas M, *et al.* (1996). *J. Biol. Chem.* **271**:27456-27461.
- 5 Rosenfeld MA, *et al.* (1992). *Cell* **68**:143-155.
- Ruano G and Kidd KK (1989). *Nucl. Acids Res.* **17**:8392.
- Russell D and Hirata R (1998). *Nature Genetics* **18**:323-328.
- Sambrook J, *et al.* (1989). Molecular Cloning: A Laboratory Manual, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- 10 Scharf SJ (1986). *Science* **233**:1076-1078.
- Schneider G, *et al.* (1998). *Nature Genetics* **18**:180-183.
- Scopes R (1982). Protein Purification: Principles and Practice, (Springer-Verlag, NY).
- Shastri *et al.* (1995). *Experientia* **51**:1028-1039.
- Shastri *et al.* (1998). *Mol. Cell. Biochem.* **181**:163-179.
- 15 Sheffield VC, *et al.* (1989). *Proc. Natl. Acad. Sci. USA* **86**:232-236.
- Sheffield VC, *et al.* (1991). *Am. J. Hum. Genet.* **49**:699-706.
- Shenk TE, *et al.* (1975). *Proc. Natl. Acad. Sci. USA* **72**:989-993.
- Shields PB (1997). *Proc. Dept. Defense BCRP Era of Hope meeting*, Vol. 1 ("Frontiers in Prevention and Detection"), pp.9-10.
- 20 Shimada T, *et al.* (1991). *J. Clin. Invest.* **88**:1043-1047.
- Shinkai Y, *et al.* (1992). *Cell* **68**:855-867.
- Shoemaker DD, *et al.* (1996). *Nature Genetics* **14**:450-456.
- Smith JR, *et al.* (1996). *Science* **274**:1371-1374.
- Smith SW, *et al.* (1994). *CABIOS* **10**:671-675.
- 25 Snouwaert JN, *et al.* (1992). *Science* **257**:1083-1088.
- Sorge J, *et al.* (1984). *Mol. Cell. Biol.* **4**:1730-1737.
- Spargo CA, *et al.* (1996). *Mol. Cell. Probes* **10**:247-256.
- Steinberg GD, *et al.* (1990). *Prostate* **17**:337-347.
- Stewart MJ, *et al.* (1992). *Hum. Gene Ther.* **3**:267-275.
- 30 Stratford-Perricaudet LD, *et al.* (1990). *Hum. Gene Ther.* **1**:241-256.
- Thierry-Mieg D, *et al.* (1995). Ace.mbl. A graphic interactive program to support shotgun and directed sequencing projects.
- Thompson S, *et al.* (1989). *Cell* **56**:313-321.
- Valancius V and Smithies O (1991). *Mol. Cell Biol.* **11**:1402-1408.
- 35 Van der Putten H, *et al.* (1985). *Proc. Natl. Acad. Sci. USA* **82**:6148-6152.

- Wagner E, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3410-3414.
- Wagner E, *et al.* (1991). *Proc. Natl. Acad. Sci. USA* **88**:4255-4259.
- Walker GT, *et al.* (1992). *Nucl. Acids Res.* **20**:1691-1696.
- Wang CY and Huang L (1989). *Biochemistry* **28**:9508-9514.
- 5 Wartell RM, *et al.* (1990). *Nucl. Acids Res.* **18**:2699-2705.
- Wells JA (1991). *Methods in Enzymol.* **202**:390-411.
- Wetmur JG and Davidson N (1968). *J. Mol. Biol.* **31**:349-370.
- White MB, *et al.* (1992). *Genomics* **12**:301-306.
- White R and Lalouel JM (1988). *Annu. Rev. Genet.* **22**:259-279.
- 10 Wilkinson GW and Akrigg A (1992). *Nucleic Acids Res.* **20**:2233-2239.
- Wolff JA, *et al.* (1990). *Science* **247**:1465-1468.
- Wolff JA, *et al.* (1991). *BioTechniques* **11**:474-485.
- Woolf CM (1960). *Cancer* **13**:739-744.
- Wu DY and Wallace RB (1989). *Genomics* **4**:560-569.
- 15 Wu CH, *et al.* (1989). *J. Biol. Chem.* **264**:16985-16987.
- Wu GY, *et al.* (1991). *J. Biol. Chem.* **266**:14338-14342.
- Xu J, *et al.* (1998). *Nat. Genet.* **20**:175-179.
- Zenke M, *et al.* (1990). *Proc. Natl. Acad. Sci. USA* **87**:3655-3659.
- U.S. Patent No. 3,817,837
- 20 U.S. Patent No. 3,850,752
- U.S. Patent No. 3,939,350
- U.S. Patent No. 3,996,345
- U.S. Patent No. 4,275,149
- U.S. Patent No. 4,277,437
- 25 U.S. Patent No. 4,366,241
- U.S. Patent No. 4,376,110
- U.S. Patent No. 4,486,530
- U.S. Patent No. 4,554,101
- U.S. Patent No. 4,683,195
- 30 U.S. Patent No. 4,683,202
- U.S. Patent No. 4,816,567
- U.S. Patent No. 4,868,105
- U.S. Patent No. 4,873,191
- U.S. Patent No. 5,252,479
- 35 U.S. Patent No. 5,270,184

U.S. Patent No. 5,409,818

U.S. Patent No. 5,436,146

U.S. Patent No. 5,455,166

U.S. Patent No. 5,550,050

5 U.S. Patent No. 5,691,198

U.S. Patent No. 5,735,500

U.S. Patent No. 5,747,469

Hitzeman *et al.*, EP 73,675A

EPO Publication No. 225,807

10 EP 425,731A

European Patent Application Publication No. 0332435

WO 84/03564

WO 90/07936

WO 92/19195

15 WO 93/07282

WO 94/25503

WO 95/01203

WO 95/05452

WO 96/02286

20 WO 96/02646

WO 96/11698

WO 96/40871

WO 96/40959

WO 97/12635

WHAT IS CLAIMED IS:

1. An isolated nucleic acid coding for an HPC2 polypeptide, said polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2 or a modified form which is functionally equivalent.
2. The isolated nucleic acid of claim 1 wherein said DNA has the nucleotide sequence (a) set forth in SEQ ID NO:1, (b) its complement, (c) a corresponding RNA or (d) a nucleotide sequence which hybridizes under stringent conditions to the nucleotide sequence of (a), (b) or (c).
3. The isolated nucleic acid of claim 1 wherein said DNA has the nucleotide sequence set forth in SEQ ID NO:3, SEQ ID NO:28, or complements thereof.
4. The isolated nucleic acid of claim 1 which is a DNA comprising an allelic variant of the nucleotide sequence set forth in SEQ ID NO:1, its complement or a corresponding RNA.
5. The isolated nucleic acid of claim 1 coding for a mutated form of the HPC2 polypeptide set forth in SEQ ID NO:2.
6. The isolated nucleic acid of claim 5 which is a DNA comprising a mutated form of the nucleotide sequence set forth in SEQ ID NO:1, its complement or a corresponding RNA.
7. The isolated nucleic acid of claim 6 wherein the mutation is selected from the group consisting of a deletion mutation, a nonsense mutation, an insertion mutation, a frameshift mutation and a missense mutation.
8. An isolated nucleic acid having a nucleotide sequence selected from the group consisting of (a) SEQ ID NOs:4-27, (b) complements thereof, (c) RNAs corresponding thereto and (d) nucleic acids which hybridize under stringent conditions to the nucleotide sequences of (a), (b) or (c).

9. An isolated nucleic acid having at least 15 contiguous nucleotides of a nucleic acid as claimed in claim 1 wherein the nucleic acid sequence suitable for use as a hybridization probe to detect in a sample (i) a DNA having a nucleotide sequence selected from the nucleotide sequence set forth in SEQ ID NO:1, allelic variants thereof and mutated forms thereof, (ii) an RNA corresponding to said DNA or (iii) a nucleotide sequence which hybridizes under stringent conditions to the nucleotide sequence of (i) or (ii).
10. A set of nucleic acid probes for use in a microchip assay wherein each of said nucleic acid probes comprises at least 8 contiguous nucleotides of a nucleic acid as claimed in claim 1 and said set encompasses part or all of said nucleic acid.
11. A vector which comprises an isolated nucleic acid as claimed in claim 1.
12. An expression vector which comprises an isolated nucleic acid of claim 1 wherein the coding sequence for the HPC2 polypeptide or modified form thereof is operably linked to suitable control sequences capable of directing expression of said coding sequence in host cells for said vector.
13. Host cells transformed with a vector as claimed in claim 11.
14. A method of producing a polypeptide which is the HPC2 polypeptide of claim 1 which comprises (i) culturing the host cells containing an expression vector encoding said polypeptide under conditions suitable for the production of said HPC2 polypeptide and (ii) recovering said polypeptide.
15. A method as claimed in claim 14 which further comprises labeling the recovered polypeptide.
16. A preparation of human HPC2 polypeptide substantially free of other human proteins, said polypeptide having the amino acid sequence set forth in SEQ ID NO:2.
17. A preparation of human HPC2 polypeptide substantially free of other human proteins, the amino acid sequence of said polypeptide having substantial sequence homology with

the wild-type *HPC2* polypeptide having the amino acid sequence set forth in SEQ ID NO:2, and said polypeptide having substantially similar function as the wild-type *HPC2* polypeptide.

18. A preparation of a polypeptide substantially free of other proteins, said polypeptide being a mutated human *HPC2* polypeptide obtainable by expression of a mutated form of the nucleic acid sequence set forth in SEQ ID NO:2.
19. A preparation as claimed in claim 16 wherein said polypeptide is labeled.
20. An antibody capable of specifically binding one or more polypeptides as claimed in claim 16.
21. An antigenic fragment of a polypeptide as defined in claim 16 suitable for use as an immunogen to obtain an antibody.
22. A polypeptide as defined in claim 16 in the form of a fusion protein.
23. Use of a polypeptide as defined in claim 16 as an immunogen for antibody production.
24. A use as claimed in claim 23 wherein one or more antibodies products are subsequently labeled or bound to a solid support.
25. A pair of single-stranded oligonucleotide primers for determination of a nucleotide sequence of a *HPC2* gene by a nucleic acid amplification reaction, the sequence of said primers being derived from genomic clones for *HPC2*, and the use of said primers in a nucleic acid amplification reaction resulting in the synthesis of DNA or RNA corresponding to all or part of the sequence of the *HPC2* gene.
26. A pair of primers as claimed in claim 25 for determination of all or part of the sequence of the *HPC2* gene having the nucleotide sequence set forth in SEQ ID NO:1, allelic variant or a mutated form thereof.

27. A method for identifying a mutant *HPC2* nucleotide sequence in a suspected mutant *HPC2* allele which comprises comparing the nucleotide sequence of the suspected mutant *HPC2* allele with a wild-type *HPC2* nucleotide sequence, wherein a difference between the suspected mutant and the wild-type sequence identifies a mutant *HPC2* nucleotide sequence.
28. A method for detecting an alteration in *HPC2* wherein said alteration is associated with cancer in a human, wherein if said alteration is in germline it is associated with predisposition to said cancer and if said alteration is in somatic tissue it indicates that said somatic tissue is cancerous, wherein said method comprises analyzing an *HPC2* gene or an *HPC2* gene expression product from a tissue of said human.
29. A method as claimed in claim 28 wherein the sequence of the *HPC2* gene in said sample is compared with the sequence of one or more wild-type *HPC2* gene sequences selected from the sequence set forth in SEQ ID NO:1 and wild-type allelic variants thereof.
30. The method of claim 28 wherein said expression product is selected from the group consisting of mRNA of the *HPC2* gene and a *HPC2* polypeptide encoded by the *HPC2* gene.
31. The method of claim 28 wherein one or more of the following procedures is carried out:
- (a) observing shifts in electrophoretic mobility of single-stranded DNA from said sample on non-denaturing polyacrylamide gels;
 - (b) hybridizing a *HPC2* gene probe to genomic DNA isolated from said sample under conditions suitable for hybridization of said probe to said gene;
 - (c) determining hybridization of an allele-specific probe to genomic DNA from said sample;
 - (d) amplifying all or part of the *HPC2* gene from said sample to produce an amplified sequence and sequencing the amplified sequence;
 - (e) determining by nucleic acid amplification the presence of a specific *HPC2* mutant allele in said sample;
 - (f) molecularly cloning all or part of the *HPC2* gene from said sample to produce a cloned sequence and sequencing the cloned sequence;

(g) determining whether there is a mismatch between molecules (1) *HPC2* gene genomic DNA or *HPC2* mRNA isolated from said sample, and (2) a nucleic acid probe complementary to the human wild-type *HPC2* gene DNA, when molecules (1) and (2) are hybridized to each other to form a duplex;

5 (h) amplification of *HPC2* gene sequences in said sample and hybridization of the amplified sequences to nucleic acid probes which comprise wild-type *HPC2* gene sequences;

10 (i) amplification of *HPC2* gene sequences in said tissue and hybridization of the amplified sequences to nucleic acid probes which comprise mutant *HPC2* gene sequences;

(j) screening for a deletion mutation;

(k) screening for a point mutation;

(l) screening for an insertion mutation;

15 (m) determining *in situ* hybridization of the *HPC2* gene in said sample with one or more nucleic acid probes which comprise the *HPC2* gene sequence or a mutant *HPC2* gene sequence;

(n) immunoblotting;

(o) immunocytochemistry;

20 (p) assaying for binding interactions between *HPC2* protein isolated from said tissue and a binding partner capable of specifically binding the polypeptide expression product of a *HPC2* mutant allele and/or a binding partner for the *HPC2* polypeptide having the amino acid sequence set forth in SEQ ID NO:3; and

(q) assaying for the inhibition of biochemical activity of said binding partner.

25 32. A non-human animal which carries an altered *HPC2* allele in its genome.

33. A non-human animal in which its native *HPC2* alleles have been replaced by *HPC2* alleles from a second animal.

30 34. A non-human animal in which its native *HPC2* alleles have been disrupted or deleted.

35. A non-human animal in which one or both of its native *HPC2* alleles have been (a) modified to contain a DNA system to conditionally knockout said alleles or (b) have

been replaced by *HPC2* alleles from a second animal modified to contain a DNA system to conditionally knockout said alleles.

36. A cell line isolated from the non-human animal of claim 32.

37. An isolated mutant *HPC2* which cannot form a complex with a wild-type binding partner with which wild-type *HPC2* does form a complex.

38. An isolated protein complex comprising *HPC2* and its binding partner.

39. A protein complex comprising a fragment of *HPC2* and a fragment of its binding partner.

40. An isolated antibody immunoreactive with the protein complex of claim 38.

41. The antibody of claim 40 wherein said antibody is not immunoreactive with either pure *HPC2* or pure *HPC2*-binding partner.

42. The antibody of claim 40 wherein said antibody is a monoclonal antibody.

43. A method for supplying a wild-type *HPC2* gene function or a *HPC2* function substantially similar to said wild-type function to a cell which has lost said gene function or has altered gene function by virtue of a mutation in said *HPC2* gene, wherein said method comprises introducing into the cell a nucleic acid which restores an *HPC2* function in said cell, said nucleic acid selected from the group consisting of a wild-type *HPC2* gene nucleic acid or a nucleic acid substantially homologous to said wild-type *HPC2* gene nucleic acid, such that said nucleic acid is expressed in said cell.

44. The method of claim 43 wherein said nucleic acid is a portion of wild-type *HPC2* gene, said portion encoding a part of said wild-type gene polypeptide which is biologically functional.

45. A method for supplying a wild-type *HPC2* gene function or a *HPC2* function substantially similar to wild-type to a cell which has lost said gene function or has altered

gene function by virtue of a mutation in said *HPC2* gene, wherein said method comprises introducing into said cell a molecule which restores an *HPC2* function in said cell, said molecule selected from the group consisting of all or a part of a wild-type *HPC2* polypeptide which is biologically functional, a polypeptide substantially homologous to said wild-type *HPC2* polypeptide and a molecule which mimics the function of said wild-type *HPC2* polypeptide.

46. A method for diagnosing a predisposition for cancer in a human wherein said method comprises assaying for the ability of *HPC2* or a fragment of *HPC2* from said human to form a complex with a binding partner to which wild-type *HPC2* binds wherein an inability to form said complex is indicative of a predisposition to cancer.
47. The method of claim 46 wherein said assay comprises a yeast two-hybrid assay.
48. The method of claim 46 wherein said assay comprises measuring *in vitro* a complex formed by mixing said binding partner and *HPC2* purified from said human.
49. The method of claim 46 wherein said assay comprises measuring *in vitro* a complex formed by mixing *HPC2* and said binding partner purified from said human.
50. The method of claim 46 wherein said complex is measured by binding with an antibody specific for a *HPC2*-said binding partner complex.
51. The method of claim 46 wherein said assay comprises mixing an antibody specific for a *HPC2*-said binding partner complex with a tissue extract from said person, wherein the lack of formation of a *HPC2*-said binding partner-antibody complex between said antibody and said tissue extract is indicative of a predisposition to cancer.
52. A method for determining whether a mutation in *HPC2* is predispositive for cancer, wherein said method comprises binding a *HPC2* with said mutation to a binding partner to which wild-type *HPC2* binds and determining whether a complex forms, wherein the lack of a complex indicates said mutation is predispositive.

53. A method of screening for drug candidates useful in treating a cancer resulting from a mutation in *HPC2*, wherein said method involves mixing a mutant HPC2 with a wild-type binding partner to which wild-type HPC2 binds, in both the presence of a drug and the absence of said drug and measuring the amount of binding of said mutant HPC2 with said wild-type binding partner, wherein if the amount of said binding is greater in the presence of said drug than in the absence of said drug then said drug is a drug candidate for treating said cancer.

54. The method of claim 53 wherein said mutant HPC2 is a fusion protein and/or said wild-type protein is a fusion protein.

55. A method of screening for drug candidates useful in treating a cancer resulting from a mutation in *HPC2*, wherein said method involves mixing a mutant HPC2 with a wild-type binding partner, to which wild-type HPC2 binds, in both the presence of a drug and the absence of said drug and measuring the amount of binding of said mutant HPC2 with said wild-type binding partner, wherein if the amount of said binding is less in the presence of said drug than in the absence of said drug then said drug is a drug candidate for treating said cancer.

56. The method of claim 55 wherein said wild-type HPC2 is a fusion protein and/or said wild-type binding partner is a fusion protein.

57. A method of screening for drug candidates useful in treating a cancer resulting from a mutation in *HPC2*, wherein said method comprises treating an animal which is homozygous for *HPC2* containing said mutation with a drug wherein if said animal does not develop cancer said drug is a drug candidate for treating said cancer.

58. A method of screening for drug candidates useful in treating a cancer resulting from a mutation in *HPC2*, wherein said method comprises treating an animal which has a tumor and which is homozygous for *HPC2* containing said mutation with a drug wherein if said tumor regresses said drug is a drug candidate for treating said cancer.

59. The method of claim 57 wherein said animal is transgenic for *HPC2* with said mutation.

60. A method of screening for drug candidates useful in treating a cancer resulting from a mutation in *HPC2*, wherein said method comprises the steps of:

5 (a) growing a cell culture of cells which are homozygous for *HPC2* containing said mutation in the presence of a drug,

(b) growing a cell culture of cells which contain a wild-type *HPC2* gene, and

(c) growing a cell culture of cells which are homozygous for *HPC2* containing said mutation in the absence of said drug,

10 wherein if the cells in step (a) behave more like the cells in step (b) than like the cells in step (c) then said drug is a drug candidate for treating said cancer.

ABSTRACT

The present invention relates generally to the field of human genetics. Specifically, the present invention relates to methods and materials used to isolate and detect a human prostate cancer predisposing gene (HPC2), some alleles of which cause susceptibility to cancer, in particular prostate cancer. More specifically, the present invention relates to germline mutations in the HPC2 gene and their use in the diagnosis of predisposition to prostate cancer. The invention also relates to presymptomatic therapy of individuals who carry deleterious alleles of the HPC2 gene. The invention further relates to somatic mutations in the HPC2 gene in human prostate cancer and their use in the diagnosis and prognosis of human prostate cancer.

10 Additionally, the invention relates to somatic mutations in the HPC2 gene in other human cancers and their use in the diagnosis and prognosis of human cancers. The invention also relates to the therapy of human cancers which have a mutation in the HPC2 gene, (including gene therapy, protein replacement therapy, protein mimetics, and inhibitors). The invention further relates to the screening of drugs for cancer therapy. Finally, the invention relates to the screening of the HPC2 gene for mutations, which are useful for diagnosing the predisposition to prostate cancer.

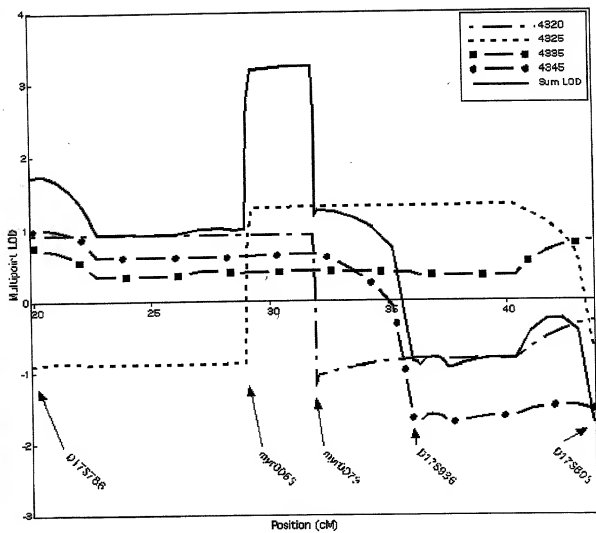
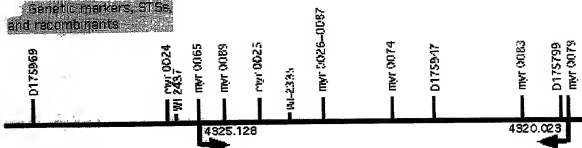
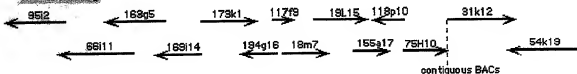


FIG. 1

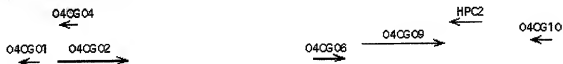
Genetic markers, STSs,
and recombinants



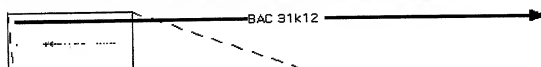
BAC contig



Transcription units



BAC 31k12 with 2 exons of O4CG09 and the HPC2 transcription unit



A 40 kb segment of 31k12 with 2 exons of O4CG09 and the HPC2 transcription unit

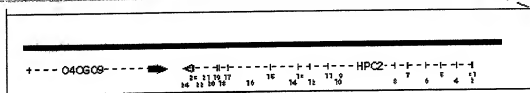


FIG. 2A

FIG. 2B

Hs.HPC2.exon1 cggggcgTaaagtaccggcggtttctcagttttgtggagacggggcg
 Hs.HPC2.exon1.pep
 Mm.HPC2.exon2 tggcgggcgTaaagggtctggctgctctcagcctgtgtggtcgggtgc
 Mm.HPC2.exon1.pep
 |-----|-----|-----|-----|-----|
 1 11 21 31 41

Hs.HPC2.exon1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTGCGGCGGCGGAGCGACCATGTC
 Hs.HPC2.exon1.pep M--W--A--L--C--S--L--L--R--S--A--A--G--R--T--M--S--
 Mm.HPC2.exon2 ATGTGGGCGCTCGCTCACTGTTGCGTCCCTTGGCTGCGCACCATGTC
 Mm.HPC2.exon1.pep M--W--A--L--R--S--L--L--R--P--L--G--L--R--T--M--S--
 |-----|-----|-----|-----|-----|
 51 61 71 81 91

Hs.HPC2.exon1 GCAGGGACGACCATATCGCAGGACCCGCGCGCGAGCGCGCGGCA
 Hs.HPC2.exon1.pep -Q--G--R--T--I--S--Q--A--P--A--R--R--E--R--P--R--K
 Mm.HPC2.exon2 GCAGGGT-----T-CGCTGCTCGGCGCGCGCACCCA
 Mm.HPC2.exon1.pep -Q--G-----S--A--R--R--P--R--P--R--K
 |-----|-----|-----|-----|-----|
 101 111 121 131 141

Hs.HPC2.exon1 AGGACCCGCTCGGCGACCTGCGCACGCGAGAGAAGCGGGACCGTCGGGG
 Hs.HPC2.exon1.pep --D--P--L--R--H--L--R--T--R--E--R--R--G--P--S--G--
 Mm.HPC2.exon2 AAGACCCACTGCGACACCTGCGTACGCGGGAGAAAGCGCGGCCGGGT--
 Mm.HPC2.exon1.pep --D--P--L--R--H--L--R--T--R--E--R--R--G--P--G--
 |-----|-----|-----|-----|-----|
 151 161 171 181 191

Hs.HPC2.exon1 TGCTCGGCGGCCCCAACACCGTGTACCTGCAGGTGTTGGCAGCGGGTAG
 Hs.HPC2.exon1.pep C--S--G--G--P--N--T--V--Y--L--Q--V--V--A--A--G--S--
 Mm.HPC2.exon2 ---CCCGGGGCCCCGAACCGTGTACCTGCAGGTGTTGGCAGCGGGCGG
 Mm.HPC2.exon1.pep --P--G--G--P--N--T--V--Y--L--Q--V--V--A--A--G--G--
 |-----|-----|-----|-----|-----|
 201 211 221 231 241

Hs.HPC2.exon1 CCGGGACTCGGGGCGCGGCTCTACGTCTTCTCCGAGTTCAACCGgtcag
 Hs.HPC2.exon1.pep -R--D--S--G--A--A--L--Y--V--F--S--E--F--N
 Mm.HPC2.exon2 CCGGGACGCGGGGGTCTCTCTATGTCTTCTCGGAATACAACAggtcag
 Mm.HPC2.exon1.pep -R--D--A--G--A--A--L--Y--V--F--S--E--Y--N
 |-----|-----|-----|-----|-----|
 251 261 271 281 291

Hs.HPC2.exon1 tcaacgagccacgccccgtcccgctggggccctcagtcggcgagcctct
 Hs.HPC2.exon1.pep
 Mm.HPC2.exon2 agtgggcccagcagccctgggggatggcccccagcgccagctgctcgggag
 Mm.HPC2.exon1.pep
 |-----|-----|-----|-----|-----|
 301 311 321 331 341

Figure 3

SEQUENCE LISTING

<110> Tavtigian, Sean V.
Teng, David H.F.
Simard, Jacques
Rommens, Johanna M.
Myriad Genetics, Inc.

<120> Chromosome 17q-Linked Prostate Cancer Susceptibility
Gene

<130> HPC2 Gene

<140>

<141>

<150> US 60/107,468

<151> 1999-11-06

<160> 213

<170> PatentIn Ver. 2.0

<210> 1

<211> 2481

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(2478)

<400> 1

atg	tgg	gcg	ctt	tgc	tgc	ctg	ctg	cg	tcc	gcg	gcc	gga	cgc	acc	atg	48
Met	Trp	Ala	Leu	Cys	Ser	Leu	Leu	Arg	Ser	Ala	Ala	Gly	Arg	Thr	Met	
1				5					10					15		

tgc	cag	gga	cgc	acc	ata	tgc	cag	gca	ccc	gcc	cgc	cgc	gag	cgg	ccg	96
Ser	Gln	Gly	Arg	Thr	Ile	Ser	Gln	Ala	Pro	Ala	Arg	Arg	Glu	Arg	Pro	
			20					25					30			

cgc	aag	gac	ccg	ctg	cgg	cac	ctg	cgc	acg	cga	gag	aag	cgc	gga	ccg	144
Arg	Lys	Asp	Pro	Leu	Arg	His	Leu	Arg	Thr	Arg	Glu	Lys	Arg	Gly	Pro	
		35					40					45				

tgc	ggg	tgc	tcc	ggc	ggc	cca	aac	acc	gtg	tac	ctg	cag	gtg	gtg	gca	192
Ser	Gly	Cys	Ser	Gly	Gly	Pro	Asn	Thr	Val	Tyr	Leu	Gln	Val	Val	Ala	
	50					55					60					

gcg	ggt	agc	cgg	gac	tgc	ggc	gcc	gcg	ctc	tac	gtc	ttc	tcc	gag	ttc	240
Ala	Gly	Ser	Arg	Asp	Ser	Gly	Ala	Ala	Leu	Tyr	Val	Phe	Ser	Glu	Phe	
	65				70				75					80		

aac	cgg	tat	ctc	ttc	aac	tgt	gga	gaa	ggc	gtt	cag	aga	ctc	atg	cag	288
Asn	Arg	Tyr	Leu	Phe	Asn	Cys	Gly	Glu	Gly	Val	Gln	Arg	Leu	Met	Gln	
			85					90						95		

gag	cac	aag	tta	aag	gtt	gct	cgc	ctg	gac	aac	ata	ttc	ctg	aca	cga	336
Glu	His	Lys	Leu	Lys	Val	Ala	Arg	Leu	Asp	Asn	Ile	Phe	Leu	Thr	Arg	
			100					105					110			

atg cac tgg tct aat gtt ggg ggc tta agt gga atg att ctt act tta	384
Met His Trp Ser Asn Val Gly Gly Leu Ser Gly Met Ile Leu Thr Leu	
115 120 125	
aag gaa acc ggg ctt cca aag tgt gta ctt tct gga cct cca caa ctg	432
Lys Glu Thr Gly Leu Pro Lys Cys Val Leu Ser Gly Pro Pro Gln Leu	
130 135 140	
gaa aaa tac ctc gaa gca atc aaa ata ttt tct ggt cca ttg aaa gga	480
Glu Lys Tyr Leu Glu Ala Ile Lys Ile Phe Ser Gly Pro Leu Lys Gly	
145 150 155 160	
ata gaa ctg gct gtg cgg ccc cac tct gcc cca gaa tac gag gat gaa	528
Ile Glu Leu Ala Val Arg Pro His Ser Ala Pro Glu Tyr Glu Asp Glu	
165 170 175	
acc atg aca gtt tac cag atc ccc ata cac agt gaa cag agg agg gga	576
Thr Met Thr Val Tyr Gln Ile Pro Ile His Ser Glu Gln Arg Arg Gly	
180 185 190	
aag cac caa cca tgg cag agt cca gaa agg cct ctc agc agg ctc agt	624
Lys His Gln Pro Trp Gln Ser Pro Glu Arg Pro Leu Ser Arg Leu Ser	
195 200 205	
cca gag cga tct tca gac tcc gag tgc aat gaa aat gag cca cac ctt	672
Pro Glu Arg Ser Ser Asp Ser Glu Ser Asn Glu Asn Glu Pro His Leu	
210 215 220	
cca cat ggt gtt agc cag aga aga ggg gtc agg gac tct tcc ctg gtc	720
Pro His Gly Val Ser Gln Arg Arg Gly Val Arg Asp Ser Ser Leu Val	
225 230 235 240	
gta gct ttc atc tgt aag ctt cac tta aag aga gga aac ttc ttg gtg	768
Val Ala Phe Ile Cys Lys Leu His Leu Lys Arg Gly Asn Phe Leu Val	
245 250 255	
ctc aaa gca aag gag atg ggc ctc cca gtt ggg aca gct gcc atc gct	816
Leu Lys Ala Lys Glu Met Gly Leu Pro Val Gly Thr Ala Ala Ile Ala	
260 265 270	
ccc atc att gct gct gtc aag gac ggg aaa agc atc act cat gaa gga	864
Pro Ile Ile Ala Ala Val Lys Asp Gly Lys Ser Ile Thr His Glu Gly	
275 280 285	
aga gag att ttg gct gaa gag ctg tgt act cct cca gat cct ggt gct	912
Arg Glu Ile Leu Ala Glu Leu Cys Thr Pro Pro Asp Pro Gly Ala	
290 295 300	
gct ttt gtg gtg gta gaa tgt cca gat gaa agc ttc att caa ccc atc	960
Ala Phe Val Val Val Glu Cys Pro Asp Glu Ser Phe Ile Gln Pro Ile	
305 310 315 320	
tgt gag aat gcc acc ttt cag agg tac caa gga aag gca gat gcc ccc	1008
Cys Glu Asn Ala Thr Phe Gln Arg Tyr Gln Gly Lys Ala Asp Ala Pro	
325 330 335	
gtg gcc ttg gtg gtt cac atg gcc cca gca tct gtg ctt gtg gac agc	1056
Val Ala Leu Val Val His Met Ala Pro Ala Ser Val Leu Val Asp Ser	
340 345 350	

agg tac cag cag tgg atg gag agg ttt ggg cct gac acc cag cac ttg	1104
Arg Tyr Gln Gln Trp Met Glu Arg Phe Gly Pro Asp Thr Gln His Leu	
355 360 365	
gtc ctg aat gag aac tgt gcc tca gtt cac aac ctt cgc agc cac aag	1152
Val Leu Asn Glu Asn Cys Ala Ser Val His Asn Leu Arg Ser His Lys	
370 375 380	
att caa acc cag ctc aac ctc atc cac cgg gac atc ttc ccc etg ctc	1200
Ile Gln Thr Gln Leu Asn Leu Ile His Pro Asp Ile Phe Pro Leu Leu	
385 390 395 400	
acc agt ttc cgc tgt aag aag gag ggc ccc acc ctc agt gtg ccc atg	1248
Thr Ser Phe Arg Cys Lys Lys Glu Gly Pro Thr Leu Ser Val Pro Met	
405 410 415	
gtt cag ggt gaa tgc ctc ctc aag tac cag ctc cgt ccc agg agg gag	1296
Val Gln Gly Glu Cys Leu Leu Lys Tyr Gln Leu Arg Pro Arg Arg Glu	
420 425 430	
tgg cag agg gat gcc att att act tgc aat cct gag gaa ttc ata gtt	1344
Trp Gln Arg Asp Ala Ile Ile Thr Cys Asn Pro Glu Glu Phe Ile Val	
435 440 445	
gag gcg ctg cag ctt ccc aac ttc cag cag agc gtg cag gag tac agg	1392
Glu Ala Leu Gln Leu Pro Asn Phe Gln Gln Ser Val Gln Glu Tyr Arg	
450 455 460	
agg agt gcg cag gac ggc cca gcc cca gca gag aaa aga agt cag tac	1440
Arg Ser Ala Gln Asp Gly Pro Ala Pro Ala Glu Lys Arg Ser Gln Tyr	
465 470 475 480	
cca gaa atc atc ttc ctt gga aca ggg tct gcc atc ccg atg aag att	1488
Pro Glu Ile Ile Phe Leu Gly Thr Gly Ser Ala Ile Pro Met Lys Ile	
485 490 495	
cga aat gtc agt gcc aca ctt gtc aac ata agc ccc gac acg tct ctg	1536
Arg Asn Val Ser Ala Thr Leu Val Asn Ile Ser Pro Asp Thr Ser Leu	
500 505 510	
cta ctg gac tgt ggt gag ggc aca ttt ggg cag ctg tgc cgt cat tac	1584
Leu Leu Asp Cys Gly Glu Gly Thr Phe Gly Gln Leu Cys Arg His Tyr	
515 520 525	
gga gac cag gtg gac agg gtc ctg ggc acc ctg gct gct gtg ttt gtg	1632
Gly Asp Gln Val Asp Arg Val Leu Gly Thr Leu Ala Ala Val Phe Val	
530 535 540	
tcc cac ctg cac gca gat cac cac acg ggc ttg cca agt atc ttg ctg	1680
Ser His Leu His Ala Asp His His Thr Gly Leu Pro Ser Ile Leu Leu	
545 550 555 560	
cag aga gaa cgc gcc ttg gca tct ttg gga aag ccg ctt cac cct ttg	1728
Gln Arg Glu Arg Ala Leu Ala Ser Leu Gly Lys Pro Leu His Pro Leu	
565 570 575	
ctg gtg gtt gcc ccc aac cag ctc aaa gcc tgg ctc cag cag tac cac	1776
Leu Val Val Ala Pro Asn Gln Leu Lys Ala Trp Leu Gln Gln Tyr His	
580 585 590	

aac cag tgc cag gag gtc ctg cac cac atc agt atg att cct gcc aaa	1824
Asn Gln Cys Gln Glu Val Leu His His Ile Ser Met Ile Pro Ala Lys	
595 600 605	
tgc ctt cag gaa ggg gct gag atc tcc agt cct gca gtg gaa aga ttg	1872
Cys Leu Gln Glu Gly Ala Glu Ile Ser Ser Pro Ala Val Glu Arg Leu	
610 615 620	
atc agt tgc ctg ttg cga aca tgt gat ttg gaa gag ttt cag acc tgt	1920
Ile Ser Ser Leu Leu Arg Thr Cys Asp Leu Glu Glu Phe Gln Thr Cys	
625 630 635 640	
ctg gtg cgg cac tgc aag cat gcg ttt ggc tgt gcg ctg gtg cac acc	1968
Leu Val Arg His Cys Lys His Ala Phe Gly Cys Ala Leu Val His Thr	
645 650 655	
tct ggc tgg aaa gtg gtc tat tcc ggg gac acc atg ccc tgc gag gct	2016
Ser Gly Trp Lys Val Val Tyr Ser Gly Asp Thr Met Pro Cys Glu Ala	
660 665 670	
ctg gtc cgg atg ggg aaa gat gcc acc ctc ctg ata cat gaa gcc acc	2064
Leu Val Arg Met Gly Lys Asp Ala Thr Leu Leu Ile His Glu Ala Thr	
675 680 685	
ctg gaa gat ggt ttg gaa gag gaa gca gtg gaa aag aca cac agc aca	2112
Leu Glu Asp Gly Leu Glu Glu Glu Ala Val Glu Lys Thr His Ser Thr	
690 695 700	
acg tcc caa gcc atc agc gtg ggg atg cgg atg aac gcg gag ttc att	2160
Thr Ser Ser Gln Ala Ile Ser Val Gly Met Arg Met Asn Ala Glu Phe Ile	
705 710 715 720	
atg ctg aac cac ttc agc cag cgc tat gcc aag gtc ccc ctc ttc agc	2208
Met Leu Asn His Phe Ser Gln Arg Tyr Ala Lys Val Pro Leu Phe Ser	
725 730 735	
ccc aac ttc agc gag aaa gtg gga gtt gcc ttt gac cac atg aag gtc	2256
Pro Asn Phe Ser Glu Lys Val Gly Val Ala Phe Asp His Met Lys Val	
740 745 750	
tgc ttt gga gac ttt cca aca atg ccc aag ctg att ccc cca ctg aaa	2304
Cys Phe Gly Asp Phe Pro Thr Met Pro Lys Leu Ile Pro Pro Leu Lys	
755 760 765	
gcc ctg ttt gct ggc gac atc gag gag atg gag gag cgc agg gag aag	2352
Ala Leu Phe Ala Gly Asp Ile Glu Glu Met Glu Glu Arg Arg Glu Lys	
770 775 780	
egg gag ctg cgg cag gtg cgg gcg gcc ctc ctg tcc agg gag ctg gca	2400
Arg Glu Leu Arg Gln Val Arg Ala Ala Leu Leu Ser Arg Glu Leu Ala	
785 790 795 800	
ggc gcc ctg gag gat ggg gag cct cag cag aag cgg gcc cac aca gag	2448
Gly Gly Leu Glu Asp Gly Glu Pro Gln Gln Lys Arg Ala His Thr Glu	
805 810 815	
gag cca cag gcc aag aag gtc aga gcc cag tga	2481
Glu Pro Gln Ala Lys Lys Val Arg Ala Gln	
820 825	

<210> 2
 <211> 826
 <212> PRT
 <213> Homo sapiens

<400> 2
 Met Trp Ala Leu Cys Ser Leu Leu Arg Ser Ala Ala Gly Arg Thr Met
 1 5 10 15
 Ser Gln Gly Arg Thr Ile Ser Gln Ala Pro Ala Arg Arg Glu Arg Pro
 20 25 30
 Arg Lys Asp Pro Leu Arg His Leu Arg Thr Arg Glu Lys Arg Gly Pro
 35 40 45
 Ser Gly Cys Ser Gly Gly Pro Asn Thr Val Tyr Leu Gln Val Val Ala
 50 55 60
 Ala Gly Ser Arg Asp Ser Gly Ala Ala Leu Tyr Val Phe Ser Glu Phe
 65 70 75 80
 Asn Arg Tyr Leu Phe Asn Cys Gly Glu Gly Val Gln Arg Leu Met Gln
 85 90 95
 Glu His Lys Leu Lys Val Ala Arg Leu Asp Asn Ile Phe Leu Thr Arg
 100 105 110
 Met His Trp Ser Asn Val Gly Gly Leu Ser Gly Met Ile Leu Thr Leu
 115 120 125
 Lys Glu Thr Gly Leu Pro Lys Cys Val Leu Ser Gly Pro Pro Gln Leu
 130 135 140
 Glu Lys Tyr Leu Glu Ala Ile Lys Ile Phe Ser Gly Pro Leu Lys Gly
 145 150 155 160
 Ile Glu Leu Ala Val Arg Pro His Ser Ala Pro Glu Tyr Glu Asp Glu
 165 170 175
 Thr Met Thr Val Tyr Gln Ile Pro Ile His Ser Glu Gln Arg Arg Gly
 180 185 190
 Lys His Gln Pro Trp Gln Ser Pro Glu Arg Pro Leu Ser Arg Leu Ser
 195 200 205
 Pro Glu Arg Ser Ser Asp Ser Glu Ser Asn Glu Asn Glu Pro His Leu
 210 215 220
 Pro His Gly Val Ser Gln Arg Arg Gly Val Arg Asp Ser Ser Leu Val
 225 230 235 240
 Val Ala Phe Ile Cys Lys Leu His Leu Lys Arg Gly Asn Phe Leu Val
 245 250 255
 Leu Lys Ala Lys Glu Met Gly Leu Pro Val Gly Thr Ala Ala Ile Ala
 260 265 270
 Pro Ile Ile Ala Ala Val Lys Asp Gly Lys Ser Ile Thr His Glu Gly

275 280 285
 Arg Glu Ile Leu Ala Glu Glu Leu Cys Thr Pro Pro Asp Pro Gly Ala
 290 295 300
 Ala Phe Val Val Val Glu Cys Pro Asp Glu Ser Phe Ile Gln Pro Ile
 305 310 315
 Cys Glu Asn Ala Thr Phe Gln Arg Tyr Gln Gly Lys Ala Asp Ala Pro
 325 330 335
 Val Ala Leu Val Val His Met Ala Pro Ala Ser Val Leu Val Asp Ser
 340 345 350
 Arg Tyr Gln Gln Trp Met Glu Arg Phe Gly Pro Asp Thr Gln His Leu
 355 360 365
 Val Leu Asn Glu Asn Cys Ala Ser Val His Asn Leu Arg Ser His Lys
 370 375 380
 Ile Gln Thr Gln Leu Asn Leu Ile His Pro Asp Ile Phe Pro Leu Leu
 385 390 395
 Thr Ser Phe Arg Cys Lys Lys Glu Gly Pro Thr Leu Ser Val Pro Met
 405 410 415
 Val Gln Gly Glu Cys Leu Leu Lys Tyr Gln Leu Arg Pro Arg Arg Glu
 420 425 430
 Trp Gln Arg Asp Ala Ile Ile Thr Cys Asn Pro Glu Glu Phe Ile Val
 435 440 445
 Glu Ala Leu Gln Leu Pro Asn Phe Gln Gln Ser Val Gln Glu Tyr Arg
 450 455 460
 Arg Ser Ala Gln Asp Gly Pro Ala Pro Ala Glu Lys Arg Ser Gln Tyr
 465 470 475 480
 Pro Glu Ile Ile Phe Leu Gly Thr Gly Ser Ala Ile Pro Met Lys Ile
 485 490 495
 Arg Asn Val Ser Ala Thr Leu Val Asn Ile Ser Pro Asp Thr Ser Leu
 500 505 510
 Leu Leu Asp Cys Gly Glu Gly Thr Phe Gly Gln Leu Cys Arg His Tyr
 515 520 525
 Gly Asp Gln Val Asp Arg Val Leu Gly Thr Leu Ala Ala Val Phe Val
 530 535 540
 Ser His Leu His Ala Asp His His Thr Gly Leu Pro Ser Ile Leu Leu
 545 550 555 560
 Gln Arg Glu Arg Ala Leu Ala Ser Leu Gly Lys Pro Leu His Pro Leu
 565 570 575
 Leu Val Val Ala Pro Asn Gln Leu Lys Ala Trp Leu Gln Gln Tyr His
 580 585 590
 Asn Gln Cys Gln Glu Val Leu His His Ile Ser Met Ile Pro Ala Lys

595 600 605
 Cys Leu Gln Glu Gly Ala Glu Ile Ser Ser Pro Ala Val Glu Arg Leu
 610 615 620
 Ile Ser Ser Leu Leu Arg Thr Cys Asp Leu Glu Phe Gln Thr Cys
 625 630 635 640
 Leu Val Arg His Cys Lys His His Ala Phe Gly Cys Ala Leu Val His Thr
 645 650 655
 Ser Gly Trp Lys Val Val Tyr Ser Gly Asp Thr Met Pro Cys Glu Ala
 660 665 670
 Leu Val Arg Met Gly Lys Asp Ala Thr Leu Leu Ile His Glu Ala Thr
 675 680 685
 Leu Glu Asp Gly Leu Glu Glu Glu Ala Val Glu Lys Thr His Ser Thr
 690 695 700
 Thr Ser Gln Ala Ile Ser Val Gly Met Arg Met Asn Ala Glu Phe Ile
 705 710 715 720
 Met Leu Asn His Phe Ser Gln Arg Tyr Ala Lys Val Pro Leu Phe Ser
 725 730 735
 Pro Asn Phe Ser Glu Lys Val Gly Val Ala Phe Asp His Met Lys Val
 740 745 750
 Cys Phe Gly Asp Phe Pro Thr Met Pro Lys Leu Ile Pro Pro Leu Lys
 755 760 765
 Ala Leu Phe Ala Gly Asp Ile Glu Glu Met Glu Glu Arg Arg Glu Lys
 770 775 780
 Arg Glu Leu Arg Gln Val Arg Ala Ala Leu Leu Ser Arg Glu Leu Ala
 785 790 795 800
 Gly Gly Leu Glu Asp Gly Glu Pro Gln Gln Lys Arg Ala His Thr Glu
 805 810 815
 Glu Pro Gln Ala Lys Lys Val Arg Ala Gln
 820 825

<210> 3

<211> 2958

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (51)..(2531)

<223> coding sequence as in SEQ ID NO:1

<400> 3

cgcgggcgta ggtgaccggc ggctttctca gttttggtgg agacggggcg atgtggggcg 60

tttgcgcgct gctgcggctc gcggccggac gcaccatgac gcagggacgc accatatcgc 120

aggcaccgc cgcgcgcgag cggccgcgca aggaaccgct gcggcacctg cgcacgcgag 180
 agaagcgcgg accgtcgggg tgctccggcg gcccaaac cgtgtacctg caggtggtgg 240
 cagcgggtag cggggactcg ggccgcgcgc tctacgtctt ctccagttc aaccgggtac 300
 tcttcaactg tggagaaggc gttcagagac tcatgcagga gcacaagtta aaggttgctc 360
 gcctggacaa catattcctg acacgaatgc actggtotaa tgttgggggc ttaagtggaa 420
 tgattottac tttaaaggaa accgggcttc caaagtgtgt actttctgga cctccacaac 480
 tggaaaaata cctcgaagca atcaaaatat tttctggtcc attgaaagga atagaactgg 540
 ctgtgcggcc ccaactctgc ccagaatacg aggatgaaac catgacagtt taccagatcc 600
 ccatacacag tgaacagagg aggggaaagc accaaccatg gcagagtcca gaaaggcctc 660
 tcagcaggct cagtccagag cgatcttcag actccagtg gaatgaaat gagccacacc 720
 tcccaactgg tgttagccag agaagagggg tcagggaact tccctggctc gtacttttca 780
 tctgtaagct tcaactaaag agaggaaact tcttggtgct caaagcaaag gagatgggcc 840
 tccagttgg gacagctgcc atgctccca tcatgtctgc tgtcaaggac gggaaaagca 900
 tcaactcatg aggaagagag attttggtcg aagagctgtg tactctcca gatcctggtg 960
 ctgcttttgt ggtggtagaa tgtccagatg aaagcttcat tcaaccatc tgtgagaatg 1020
 ccacctttca gaggtaccaa ggaaggcag atgccccctg ggccttggtg gttcacatgg 1080
 ccccgatc tcgtcttggt gacagcaggc accagcagtg gatggagagg ttggggcctg 1140
 accccagca cttggtcctg aatgagaact gtgcctcagt tcaaacctt cgcagccaca 1200
 agattcaaac ccagctcaac ctcatccacc cggacatctt cccctgctc accagtttcc 1260
 gctgtaagaa ggagggcccc accctcagtg tgcccatggt tcagggtgaa tgctctctca 1320
 agtaccagct cgtgccagg agggagtgcc agagggatgc cattattact tgcaactcctg 1380
 aggaattcat agttgaggcg ctgcagcttc ccaactcca gcagagcgtg caggagtaca 1440
 ggaggagtgc gcaggacgac ccagcccgag cagagaaaag aagtcagtac ccagaaatca 1500
 tcttcttggt aacagggtct gccatccgca tgaagattcg aaatgtcagt gccacacttg 1560
 toaacataag ccccgacacg tctctgtctc tggactgtgg tgagggcaca ttggggcagc 1620
 tgtgccgtca ttacggagac caggtggaca gggctcctgg caccctggct gctgtgtttg 1680
 tgtccacctg gcacgcagat caccacacgg gcttgccaag tatcttgctg cagagagaa 1740
 gcgccttggc atctttggga aagccgcttc accttttct ggtggttgcc ccaaccagc 1800
 tcaaaagcct gctccagcag taccacaacc agtgccagga ggtcctgcac cacatcagta 1860
 tgattcctgc caaatgcctt caggaagggg ctgagatctc cagtcctgca gtgaaagat 1920

tgatcagttc gctgttgoga acatgtgatt tggagagatt tcagacctgt ctggtgcggc 1980
 actgcaagca tgcgtttggc tgtgcgctgg tgcacaectc tggctggaaa gtggtctatt 2040
 ccgggggacac catgccctgc gaggtctctgg tccggatggg gaaagatgcc accctcctga 2100
 tacatgaagc caccctggaa gatggtttgg aagaggaagc agtggaaaaa acacacagca 2160
 caacgtccca agccatcagc gtggggatgc ggatgaacgc ggagttcatt atgctgaacc 2220
 acttcagcca gcgctatgcc aagggtcccc tcttcagccc caacttcagc gagaagtgg 2280
 gagtgcctt tgaccacatg aagggtctgt ttggagactt tccaacaatg cccaagctga 2340
 ttccccact gaaagccctg tttgtggcg acatcgagga gatggaggag cgacgggaga 2400
 agcgggagct gcggcaggtg cgggcggccc tctgtccag ggagctggca ggcggcctgg 2460
 aggatgggga gcctcagcag aagcgggccc acacagagga gccacaggcc aagaaggcca 2520
 gagccacgtg aagatctggg agaccctgaa ctacagaagc tgtgtgtctt ctgccccacg 2580
 cacgcacccg tatctgcctt cttgtctggt agaagctgaa gagcacggtc ccccaggagg 2640
 cagctcagga taggtggtat ggagctgtgc cgaggcttgg gctccacat aagcactagt 2700
 ctatagatgc ctcttaggac tggtgccctg cacagccgag gccacaggag ctgccacacg 2760
 gaagcaagca gatgaactaa ttctatttca aggcagtttt taaagaagtc ttgaaacag 2820
 acggcgccac ctttctctta atccagcaaa gtgattccct gcacaccaga gacaagcaga 2880
 gtaacaggat cagtgggtct aagtgtccga gacttaacga aaatagtatt tcagctgcaa 2940
 taaagattga gtttgcaa

2958

<210> 4
 <211> 295
 <212> DNA
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <222> (51)..(295)
 <223> exon 1

<400> 4
 cgccggcgta ggtgaccgac ggctttctca gttttgggtg agacgggcgc atgtgggcgc 60
 tttgctcgct gctgcggtcc gcggccggac gcaecatgtc gcagggacgc accatatcgc 120
 aggcaccgcg ccgcgcgcgag cggccgcgca aggacccgct gcggcacctg cgcacgcgag 180
 agaagcgcgg accgtcgggg tgctccggcg gcccaaacac cgtgtacctg caggtgtgtg 240
 cagcgggtag ccgggactcg ggccgcgcgc tctacgtctt ctccaggttc aaccg

295

<210> 5

<211> 51
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(51)
 <223> exon 2

<400> 5
 gtatctcttc aactgtggag aaggcggtca gagactcatg caggagcaca a 51

<210> 6
 <211> 71
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(71)
 <223> exon 3

<400> 6
 gttaaagggt gctcgccctgg acaacatatt cctgacacga atgcactggt ctaatgttgg 60
 gggcttaagt g 71

<210> 7
 <211> 65
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(65)
 <223> exon 4

<400> 7
 gaatgattct tacttttaaag gaaacogggc ttccaaagtg tgtactttct ggacctccac 60
 aactg 65

<210> 8
 <211> 58
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(58)
 <223> exon 5

<400> 8
 gaaaaatacc tcgaagcaat caaaatattt tctggtccat tgaaggaat agaactgg 58

<210> 9

<211> 69
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(69)
 <223> exon 6

<400> 9
 ctgtgcggcc ccaactctgcc ccagaatacg aggatgaaac catgacagtt taccagatcc 60
 ccatacaca 69

<210> 10
 <211> 120
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(120)
 <223> exon 7

<400> 10
 gtgaacagag gaggggaaag caccaacccat ggcagagtc agaaaggcct ctgagcaggc 60
 tcagtccaga gcgatcttca gactccgagt cgaatgaaaa tgagccacac ctccacatg 120

<210> 11
 <211> 59
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(59)
 <223> exon 8

<400> 11
 gtgttagcca gagaagaggg gtcaggggact cttccctggc cgtagcttcc atctgtaag 59

<210> 12
 <211> 59
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(59)
 <223> exon 9

<400> 12
 cttcacttaa agagaggaaa cttcttgggtg ctcaaagcaa aggagatggg cctccagtc 59

<210> 13

<211> 73
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(73)
 <223> exon 10

<400> 13
 tgggacagct gccatcgctc ccatcattgc tgctgtcaag gacgggaaaa gcatcactca 60
 tgaaggaaga gag 73

<210> 14
 <211> 113
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(113)
 <223> exon 11

<400> 14
 attttggctg aagagctgtg tactctcca gactctggtg ctgcttttgt ggtggtagaa 60
 tgtccagatg aaagcttcat tcaaccatc tgtgagaatg ccacctttca gag 113

<210> 15
 <211> 96
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(96)
 <223> exon 12

<400> 15
 gtaccaagga aaggcagatg cccccgtggc cttggtgggt cacatggccc cagcatctgt 60
 gcttgtggac agcaggtacc agcagtgatg ggagag 96

<210> 16
 <211> 139
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(139)
 <223> exon 13

<400> 16
 gtttgggcct gacaccagc acttggctcct gaatgagaac tgtgcctcag ttcacaacct 60

tcgcagccac aagattcaaa ccagctcaa cctcatccac cgggacatct tccccctgct 120
 caccagtttc cgctgtaag 139

<210> 17
 <211> 86
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(86)
 <223> exon 14

<400> 17
 aaggagggcc ccacctcag tgtgccatg gttcagggtg aatgcctcct caagtaccag 60
 ctccgtccca ggaggagtg gcagag 86

<210> 18
 <211> 119
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(119)
 <223> exon 15

<400> 18
 ggatgccatt attacttgca atcctgagga attcatagtt gaggcgctgc agcttcccaa 60
 cttccagcag agcgtgcagg agtacagag gagtgcgag gacggccag cccagcag 119

<210> 19
 <211> 97
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (1)..(97)
 <223> exon 16

<400> 19
 agaaaagaag tcagtagcca gaaatcatct tccttggaac aggggtctgcc atcccgatga 60
 agattcgaaa tgtcagtgcc acaattgtca acataag 97

<210> 20
 <211> 139
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature

<222> (1)..(139)

<223> exon 17

<400> 20

ccccgacacg tctctgtctac tggactgtgg tgagggcaca ttggggcagc tgtgccgtca 60

ttacggagac caggtggaca gggctcctggg caccctggct gctgtgtttg tgteccacct 120

gcacgcagat caccacacg

139

<210> 21

<211> 39

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(39)

<223> exon 18

<400> 21

ggcttgccaa gtatcttgct gcagagagaa cgcgccttg

39

<210> 22

<211> 110

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(110)

<223> exon 19

<400> 22

gcattcttgg gaaagccgct tcacctttg ctggtggttg cccccaacca gctcaaagcc 60

tggctccagc agtaccacaa ccagtgccag gaggtcctgc accacatcag

110

<210> 23

<211> 100

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(100)

<223> exon 20

<400> 23

tatgattcct gccaaatgcc ttcaggaagg ggctgagatc tcagtcctg cagtggaaag 60

attgatcagt tcgtgtttgc gaacatgtga ttggaagag

100

<210> 24

<211> 121

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(121)

<223> exon 21

<400> 24

tttcagacct gtctggtgcg gcactgcaag catgcgtttg gctgtgcgct ggtgcacacc 60

tctggtcgga aagtggtcta ttccggggac accatgccct gcgaggctct ggtccggatg 120

g

121

<210> 25

<211> 79

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(79)

<223> exon 22

<400> 25

ggaaagatgc caccctcctg atacatgaag ccaccctgga agatggtttg gaagaggaag 60

cagtggaaaa gacacacag

79

<210> 26

<211> 145

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(145)

<223> exon 23

<400> 26

cacaacgtcc caagccatca gcgtggggat gcggatgaac gcggagtcca ttatgctgaa 60

ccacttcagc cagcgctatg ccaaggtccc cctcttcagc cccaactcca gcgagaaagt 120

gggagttgcc ttgaccaca tgaag

145

<210> 27

<211> 655

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(228)

<223> exon 24

<220>

```

<221> polyA_signal
<222> (636)..(641)

<400> 27
gtctgctttg gagactttcc aacaatgcc aagctgattc cccactgaa agccctgttt 60
gctggcgaca tcgaggagat ggaggagcgc agggagaagc gggagctgcg gcaggtgcgg 120
gcggccctcc tgtccaggga gctggcaggc ggccctggagg atggggagcc tcagcagaag 180
cgggccacaca cagaggagcc acaggccaag aaggtcagag ccagtgaaag atctgggaga 240
ccotgaactc agaaggctgt gtgtcttctg cccacgcac gcaccgtat ctgccctect 300
tgctggtaga agctgaagag cacggtcccc caggaggcac ctcaggatag gtggtagtga 360
gctgtgcgca ggcttgggct cccacataag cactagtcta tagatgcctc ttaggactgg 420
tgccctggcac agccgcgggc caggaggctg ccacacggaa gcaagcagat gaactaat 480
catttcaagg cagtttttaa agaagtcttg gaaacagacg gcggcacctt tcctctaate 540
cagcaaaagt attccctgca caccagagac aagcagagta acaggatcag tgggtctaag 600
tgtccgagac ttaacgaaaa tagtatttca gctgcaataa agattgagtt tgcaa      655

<210> 28
<211> 26664
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<222> (910)..(13104)
<223> exon 1: 910-1154; exon 2: 1736-1786; exon 3:
      1925-1995; exon 4: 3025-3089; exon 5: 4361-4418;
      exon 6: 5582-5650; exon 7: 7075-7194; exon 8:
      8186-8244; exon 9: 12878-12936; exon 10:
      13032-13104;

<220>
<221> misc_feature
<222> (13756)..(22917)
<223> exon 11: 13756-13868; exon 12: 15283-15378; exon
      13: 16278-16416; exon 14: 16498-16583; exon 15:
      18583-18701; exon 16: 20349-20445; exon 17:
      22172-22310; exon 18: 22879-22917

<220>
<221> misc_feature
<222> (23045)..(26452)
<223> exon 19: 23045-23154; exon 20: 23795-23895; exon
      21: 23973-24093; exon 22: 24354-24432; exon 23:
      25026-25170; exon 24: 25812-26036; polyadenylation
      signal: 26447-26452

<220>
<221> variation
<222> (826)..(23879)

```

<223> s at positions 826 and 23180 is G or C; y at positions 1914, 5568, 7165, 16431, 1857 and 20486 is C or T; n at position 13128 is t or tgat; r at positions 22211 and 23879 is A or G.

<400> 28

```
tatcaggtga ctgaattcta tattctgaag taggagatac tgttattgct gttattacat 60
tttacacata agaaagctga ggctctgaga ggtcaagatc acgcagctaa caaatgagcc 120
aagactcttg ctttagagct tgtcctctat tcttgctttt ctttccaaaa aacactacaa 180
tttttgtttt gttttgtttt gttttgagac agggctctcga ggtgtcaccc aggcgtggagt 240
gcagtgggcg gatttcgact caccgcaacc tcgcctcccg cgcttaagcg attctctcgc 300
ctcagcctcc caagtagctg ggactacaag ctccgggacac cagctaaaaa tgatcaagtt 360
ctaacaatgta tgcatacga ttacaatgga aataaaatta gcaaaagcgt tatgctaagt 420
ctcaatacaa ttgatttctc cacatttaac cctcacacc actacaacca cctctaactc 480
aagctctgag ggactgacgt gcccgaggga cacagctctt atctggtgag aacaggagcg 540
ttttagcgaa actccaaact ctaggtccc gccctcccca ggaagccttt tctggcact 600
gtgcttcggg aagtcgcgcc ccaggagaaa aacagcttcc ggaaaaaatt gcggccggca 660
aaccggaaca gaactagggg cggggccgct tgagacgctc tagtattctc ctactctatg 720
gccactgtca attgacaagt ccgagcggtt aaagctcctt tctattggat gagcagcctc 780
gcgtaggcgg gaagctcggt gcacggcgcg ctgattgggt ggatcsgcca tgcggagcgg 840
ctaggtgtgt caccggaaac gcgggcgtag gtgaccggcg gctttctcag ttttggtgga 900
gacgggcgca tgtgggcgct ttgctcgctg ctgcggctcg cggccggagc caccatgtgc 960
cagggagcga ccatatcgca ggcaccgcc ccgccgagc ggccgcgcaa ggaccccgctg 1020
cggcacctcg gcacgcgaga gaagcgcgga ccgtcggggt gctccggcgg cccaacacc 1080
gtgtacctcg agtggtgtgc agcgggtage cgggactcgg gcgcccgcgt ctactcttcc 1140
tccagattca accggtcagt caacgagcca cgcctcgctc cgctgggccc tcagtgcggc 1200
gcagcctctg agcatcgggg cactcaccag ggcttcgggt tccctgcttc acacatgtgg 1260
ttcactgttg cgggggttgc tggagttagt gtgggtggga aatccgagat tctttgcata 1320
catgtgattt ctgcgatct gtgaagaact tcaggcctgg gtctgagcgt ccttttccca 1380
acccttgggc cccggccttg ctgtcagcac ttctggagct ccacctctt ccgtgcacc 1440
caaggccagt gtgtcgttgt tagcgtgtgg ggtggacaga tctggtgtgt agccggtggt 1500
ggagaaagga ctcatattgt ctagcacc ccacacacag gccccactc ctctccacct 1560
ctgctaagga gggctcaaaa ccaccagca taaatgtggc tcggtagtcc aacgtggact 1620
```

ttttaattttt ttttctttt tttttttcca gagtctacaa taaaacatct aattggtgtc 1680
 agagagtta cagaataaaa ccttctgaat gtcttgtga atgtttgtct tgtaggtatc 1740
 tcttcaactg tggagaaggc gttcagagac tcatgcagga gcacaagtga gtcagtctct 1800
 tgctttcgga gggggagtgtg attacggggc ttgaaagccg aaatgagagg ccagttgttt 1860
 tttatagcaa aagtggctct tgttctgttc atgttatcct gtttaaatgt ttttctatc 1920
 tttaggttaa ggttgctcgc ctggacaaca tattctgcac acgaatgcac tggctaatg 1980
 ttgggggctt aagtgggtgag tataattctt gcagtgtcag aggctggtgg gaagtctctg 2040
 ggatttttae cggttttacc atttttccaa gtctgggggt ggcagctact ttttttttt 2100
 ttttttttt tgtagtggc gtgatcttg ctcactgcaa cctttgcctt ctgggctcag 2160
 gtgatcccc caccctagcc tcccaaatag ctgggaccac acgtgtgtcc catcacact 2220
 ggctaatttt ttttgatgt tttgtagcga cgggggtttg ctatgtgtcc caggctggtc 2280
 tcaaaactct gcgatctcc tgtctcgcc tcccagagt ctgggattac aggcattgag 2340
 caccgcacct ggctggaat tctttttata ccagccagc cagcagcag acagagcatt 2400
 aaaagctgtg actcaggaga acagatttta atatggatac cactctctaa gtgttaccat 2460
 ccacttagtt tcttgcgttg cggggacaga gatttgtggc agtaaaactg agagtctagc 2520
 agtgggtgatt acagttaata tgtttaccgc agacgcatt ggcacattgg cagccacaca 2580
 cataccact gtccagatta cctgtcatt tatgtctatc aaccggaagg tcaggattgt 2640
 gttgcagcca aattgtgtg gcttgggtgc atggaccgga aggagtgaag tgttagacca 2700
 gtctccctc tcagggtga gactagggtg aggcacttag ggtgccagc cttcacttgc 2760
 atgattcctt acattttgca cactgggtgc ctgtctgctt caccctagt acagctcagc 2820
 ccattctaga ggcatttaaa gaatatttg tgtctgttac acctctagct ggcattcatt 2880
 ctgctctgta catcttccct ggttgtactt ccaaagctgg aaggtggaga tgtagataaa 2940
 tagttggatt agtacgggtt gctctctg ttagtgacga caggtcaaat tgatgagaga 3000
 tctgatttta tgcattcttt ttaggaatga ttcttacttt aaaggaaacc gggcttccaa 3060
 agtgtgtact ttctggacct ccacaactgg tgagtcttct ctgacacatc tttcaaaagc 3120
 aatctttctt ttgtaatat cagtaacaag aattttcctt ttgtcaaatc agtcttctgc 3180
 cctccagaga tacctggtg ttgaaacgct tcccttttca agttaaaaag acttgagtgc 3240
 tgattaacta tgtgacctg atcaagttac ttacacttct tgagctttag ttattctatc 3300
 tataagatga ctatcaggt tcatagagtt gttaaagatt aaatgacgta gcagcacata 3360
 taaagcacta aatcacttta ttagatatat gtttggcacc aagtaggcac acaagaaagg 3420

gcagcttttg tttttattca ataaatttct gacatcttct tacctttcag tccagcttat 3480
 tacactcttg agaaggcgtg tgtgtgttgt tgaatataac agttcatttt ccagtcctta 3540
 agaagaaagt caccaagacc tggttaagtct tccccaaaa taacgtttga aatccatcca 3600
 tttgtctctt attgaggcct tctctatttc tgtttttctat gctgtgaaac tacaatagcc 3660
 tcccatatc attctgcct tctgtaatc catctgocac acagcagcca gagaggctac 3720
 ttcaagacag aaaagtagtg tgtcaactgc caccctaaag cccttcattg gctccccatt 3780
 gcaatacaat caaaacacct tgatatggcc tacaagctct gtaggcccgg gcgcctaccc 3840
 acacttccat ctgtaccat cgctgaactg cagctgcatg ggctgactct tatgtccctc 3900
 taactccctg gccacttcag gactttcgcc ctcccgccgg tccctctgc ctctctaat 3960
 tgctgcctat attgttactg aaccttcagg gctcagctag agggtcattt actccagaac 4020
 tgctctctt tctctagaca agttggatcc cagcctcttg tatttttcat tttccttgca 4080
 gagcacttag cataatgcc ctaagctggt tctgttatcg tgtttcctt tgtctctccc 4140
 actggcctga ttagagcaag gccctccatct ctttttctg ctatatcctt ggcctctgat 4200
 ataatggata ctacagtaaat atttgtaata aatgatgttc aaaatatita ctaagctttg 4260
 ttttatgttg atacctattg gtaacctttt aaatacttga atagtgtctg tgttctacat 4320
 ttgttcaacc ataactgtc atttctttgt ttttcattag gaaaataacc tcgaagcaat 4380
 caaaatattt totgggtccat tgaaaggaat agaactgggt acgtctttgt ctgtgactca 4440
 tcctctgcta tttctaactt atatatgcc tgacctctca aattagaatc cattaaaaac 4500
 atcaacatca aacctcaaaa tcaaatgctt catcaccacg agattttttt tttttttttt 4560
 ttttttggat agagtcttgc tttcttacca ggctggagtg cagtggcatg atctcggtct 4620
 actgcaacct ccacctctg ggttcaagcc attctccac ctacgctcc tgagtageta 4680
 ggactacagg cgcctgccat cagctcagc taattttttg tatttttagt agagacgggg 4740
 tttccacctg ttggccagga tagtctgat ctcttgacct tgtgatctgc ccgctcagc 4800
 ctcccaaat gagctaccat gtggctggag atgggatttc taaatagtga cattttctgt 4860
 gttccacct catgctgtaa aaataggggc caggctggca ggagtgattg aacagctgat 4920
 gctgcctgt gtacatgtg tgtggcattc tccatccaga cggcagggct cctgcctcag 4980
 ttccagaggt gcttctctgc gttgagttgc tttgagttgg gggcgggggt gacaaggggt 5040
 ccctagaggt tttgtggcca accttgtaca ttgaaacgca gctccagctg cgcagggggg 5100
 cttacagcct cttgatggga agaggcctca ctgaggatgc tagtagggct cttgtccttg 5160
 cactgggtgt tatctgtggc ttgttaatac tcctctttta tagaaacact aatactttgt 5220

ttcaaaatat acatcagctc ttctggtttg cgatgatagg ttccctggct tcactattct 5280
 gtttgtaaac ttgggtctct gaaagttgag tactagtctt ttgtttttca atttttaacg 5340
 gatagtccac aaagattata atgtcttttc atctggctgt agtaaatata aatggctgac 5400
 caaaatacac ttttatttat ttccataaaa tggtaatctc cttagaaagt ctgggttttcg 5460
 tgtcagattc ccaccataat tctgaggcaa ttcagttgct cgtggttggt gatcctgaag 5520
 ttactcttcc cacacatctt cactaatgca atcactttgc tgttgygygg ttttcttgta 5580
 gctgtgcggc cccactctgc ccagaatac gaggatgaaa ccatgacagt ttaccagatc 5640
 cccatacaca gtgagtatga aagccagggt tcccaggagg aggggtgacg tcttgagtaa 5700
 agaaaaatg gatgaaaata gaaactgaac acttgctgtg ggcacccctg tttgtgtct 5760
 gagcatgatt agaaaattta gttgaggaat gaagatatg ctctgccct ggctataaaa 5820
 cttacggatg tctgacttat gcctaagtat agtgattatg ctttggaaata ttgataaatc 5880
 aagcactggt ggtaaataga ttgcattcaa gtttgcacat tcattgcttg gaggtttttt 5940
 cccacaggcg taataccctc ttttgatcag acgatcatga agaggtttgc acagatagat 6000
 ttttttaaat aaataatgat tacagcaacc taaaagaagt gttgttgggg gttagaagct 6060
 cctgcaaatt ccgaagtatc agggccagat gatgtggtct tagcttagga aaagagttag 6120
 tcttgctctt gaacttggtc aaagacattc atgtctggtt ttacttacat gtgaagagag 6180
 taccaagcag taggggtatt tcttggttag tactaactaa tgtgatgctt actaagtagt 6240
 gctgatgggt gacagaccag agcaccagc aaaggccaga gaagtcacga acctggcgag 6300
 gagatgaggc ttactctgac tgaaggcaga aggcagcagg gaggagagga atgtgccgga 6360
 gcaatggcac aagtgtctct aggcagatgc tgtgatgagc tgatcagcac tccattgcc 6420
 tggcttgctc ctctgctca gatgccttct ctacactgac cctgctgtga gccaccccca 6480
 gcctgagttg catccacctg tttgtgtctc atttccagca cctgttctt cgtccatg 6540
 catgtgacag ttaactttca tatgtgattt gcgtgatcga tgttaacatg ctgagttttg 6600
 ccgatcactg ttttttcagt gtccagcggc cctcagttag tgaacttacg ttattctcg 6660
 ttgcagctgt gcttttagct cttagagcag cgaatttttt tcccttgatc ttgagcctta 6720
 actaaatgta aaatgaggtc cttcttgtag ataggtaacc tttgggtcta tgtgttttag 6780
 cgggagtgat gataataaat aagcatgtct acaaccaca tgctgtttag ataacagctt 6840
 gttgagtttg tactgtggcc gaggtctgta gctaagcaga aacataaaca ttaataggac 6900
 ataggtgcag cccagaaacc aggtaggaag ttaactaact agttatttcc tactgtatag 6960
 taaaaggtgt gctgatttaa ttggcttctt ggcattccca tgtatgaacg tctgggctt 7020

ggctgtcagc tcacottgtg cagtgtgtaa ttgggtgga tctgtactga ccaggtgaac 7080
 agaggagggg aaagcaccaa ccatggcaga gtccagaaag gcctctcage aggcctcagtc 7140
 cagagcgatc ttcagactcc gagtygaatg aaaatgagcc acaccttcca catggaata 7200
 gtataaaca aaacagagcag cagaaaggct tgcgttttct taattctctg ccttgtaatg 7260
 cttgtagaga gtcattattg taagaaagcc aggtgtgtaa acagatcctt ctctctgggc 7320
 ttactataac ttggcccgtt gggggaatga gaagggttgt tgtaaagggt gcagcctgca 7380
 accttaataa tgaccagtc caggttttgg ccaccagggt tctgggtagg cccaaaactg 7440
 tgttctgttt tcccagagga gaacagggcc tgacaaacgg attcattttg tatttttcat 7500
 taatgtaaca ttatgcaaa ttttcatta atgtggaac tataactgct aagccaatga 7560
 gacagtcaaa tcagtgaag gctctgcacg tcttcagaa tgacagccca ctgggaaacg 7620
 gagttaaaag tccaagatga gatgtagctc aggagtcagg ccgctctggg agtttgtgtg 7680
 ccttaacaga aggtcagcgt tggcaagct cggcagctcc tcttctgtgc ctgaggtctt 7740
 gtctagtgc tgagaacagg ctgaccccta tgtgctgtcc ttgtttggat ggcacccggg 7800
 aaagactgac accagcattt tctctgcagg cctttgaact ttgtgttat ttcatatatt 7860
 atatgttta taaagccat tacaatatat tttctctgt cttctcagt cctaggtgaa 7920
 atgtgtcatt taaaaaaaaa ttcacttgcc attctaaagt tttctgggtg agagttttgt 7980
 gtttttctt tacgcaaca catctccaca taagtaggga aaaaaagctt ccttgagtat 8040
 attagtgtct tcagcctttg tattgggaca gtacgcctca ttaattttta tgtgaagtga 8100
 aattaggtat cgggtcataa tcagtctgtg atgtcttcac agctttcaca tttacctgt 8160
 gataatcaag tgtgttttct ctcagggtgt agccagagaa gaggggtcag ggactcttcc 8220
 ctggctgtag ctttcatctg taaggaaggg aagactttcc ggagggtcgt acatgactgg 8280
 ggtcttggtc agcgaactct ggtttgcaat ttttcattaa ttgagggtga ggcaactcctg 8340
 ttacctgaga caagaagaga tagcagatct tcagaaaagc tgatggaagg ccgggtgcag 8400
 tggctcacgc ctgtaatccc agcactttgg gagtccaagg cagggtgac acgaggtcag 8460
 gagtttgaga acagcctgac caacgtgggt aaacctgtc tgtactaaaa atacaaaaat 8520
 tagctgggtg tgggtggcca tgctgtaat ccagctact tgagaggcca aggcaagaga 8580
 atcgttgtaa cacaggaggg ggaggttgca gtgagttgag attgcacct tgcactccag 8640
 cctgggtgac agagcaagac tctctcaaa aaaaaaaaaa ttgatagaa atgacactgg 8700
 caatgagcct gcaacaagta ttactactga ctttccataa ttgtcatcac ttgtaggttt 8760
 cagagtttag atgctctgtt tctcaaaaa accccatact tttatttctt tttaaatttt 8820

tttccagtgc cctgtcagcc tccgtacatt tttttttttt ttttttgaga ccatgtctgt 8880
 ctccatgcgc taggctggag tgtgcagtg cacaatctcg gctcactgca gctccacct 8940
 cccaggttca agtgattctc ctgcctcagc ctcccaagta gctaggatta taggtgcgcg 9000
 ccaccacacc cagttaattt ttgtattttt agtagagatg gggtttcacc atgttggcca 9060
 ggctgggttc actcctgacc tcagggtgac caccacactt ggctcccaa aatgctggga 9120
 ttacaggcgt gaagcactgt gcttggtcca tattctttta tttttgcaa tgattgttc 9180
 ttttagaatt cagaaattat tgaaggcagc tgtgtttgtt ttcttcaac tccatcaggc 9240
 ctttattcaa agtcttttaa ctctgtttta cttttttca ttccctgca atagctaagg 9300
 tctaaccaca gattaatttg aatattagct agcattcaca aaggcctaga tctgtaactc 9360
 tgaaattggt caaatccat taaaaatttt tgttacaata agctgtttgt aagatctgac 9420
 tagtggotta ttttaatatg aattttgcat taaaaattta tcaatacaat ttgcaacaaa 9480
 tttgtctaaa tatgtgaaaa gatttcattg cctttttgtg ggcttagatt attttttaat 9540
 gttgtatttg aaatatattt ggaattgtta tctaaattct aaaagctaca agtgaataa 9600
 ataatgaaag taagtagtta atattagtgg caagatcatt gccagtatca tttctatcga 9660
 tttatttgaa taatgtgatt ttcataaaaag ttaagtacta ctgttaacag gcttattact 9720
 tgtatgttcc tgagttttag atagcaaaat cattttttta agttttaaaa atattttatt 9780
 tttgataatc tatatttata ttgtctgatt tttaaactgt tttctatggt aatcttttaa 9840
 tcgtattcct gotttccgga ataggtaaca gtgagcatga tgaaaagtga caagctcact 9900
 tttacacact cgggcagttg cctattatc aggcagcgt tcttgggggc tgccagctgc 9960
 ctgcccctggc ttttccatct ccttctctgc tgtcttctgc ggctcctct gagggctgct 10020
 gtcactggat tagcctataa cgcctttccc ctcttctaata taatttgcgt ctctcagggtg 10080
 aggttttgga aagcaataaa gctgagctag gtcaagttcc aggagctctt tggcatgagg 10140
 acctgaaaaa ctcatctgtt ggaagacctc ggctttgggc agctggtgca ctgttggggc 10200
 gttattggct gcgttctggc tctcatcagt ctccagata ctctgcattc ctgagagagg 10260
 aacatatctc catgggttga gttcagctcc caggagatg ggtttccctg ccttaagtgc 10320
 gcaagtacct tttttttct ttttttgaga cagagtctcg ctctgtcacc aggtggagt 10380
 gcagtggtgc gatcttggtc cactgcaacc tctgctccc agggtaacga agttctcctg 10440
 cctcagctc ccgagtagct gggactacag gagcgacca ccatgccag ctaatttttg 10500
 tattttttta gtagagacgg ggtttcacca tgttggccag gatggtctg atctcttgat 10560
 tctctgacc cctgctctg gctcccaaa gtgctgggat tacaggcgtg agccatcatg 10620

accagccttt atgtttcttt gtttgtttt tttttctgag atggagctctc gctctgttgc 10680
 ccaggctgga gtgcagtggt gccatctcga cttactgcaa cctctgcctt ccagggtcaa 10740
 gtgattcctt gctcagcct cccgtgtagc tgggatcaca ggtgcctgcc accatgcccg 10800
 gctaattttt gtattgttag tagacacagg gtttcgccat gttggccagg ctagtctcga 10860
 actcctgacc tcaagtgatc tgccttctc agcctcctaa agtgcctggg ttacaggagt 10920
 gaaccaccat gccagcctt caattacctt ttatttattt tttattttt 10980
 agacggagtc tttctgtgtt gccaggctg gagtgcggtg gcgcaatctt agctcactgc 11040
 aacctctccc tcccaggctc aagtgtttct catgcatcag ctcccgaggt agctgggact 11100
 tcaggtgcc gccaccacac ttggctaatt tttgtgttt tagtagagac ggggtttcac 11160
 catgttgccc aggtgtgtct tgaatttctg acctcaaat atcctctcgc ttcagcctcc 11220
 caaagtctgc ggattacagg cgtgagccac tgcccccac agcaagtacc ttttaaacat 11280
 tagagacatt tagttgccat cctcaaacc gtttgggtgt gtggagagaa tgttgggtcg 11340
 tgacatgggt gttagtattc taaagatgtc agccatcaat catcactgtg tgaatgtcac 11400
 actgaagctg taatccttca tctaggatga ttttttttaa gatggaaaat tctacaacc 11460
 tgagaataag gatttcagat ccaaatttga gactcagccc tacgagtaac tctttaactt 11520
 cagagagtta aaagaagatg cacagttgat gaagatttaa agggaaaaat ggaatcaaa 11580
 tgtcatttag cactcaaagg cctacatgtc atttctgaca ttttctgtt tgtgtgaaa 11640
 ttttttttct ctataaaatg attgtgaagt tttctggtag aattatttgt tgcctttcta 11700
 atgtaatagc atattaggtt ttttttttct tcttttttct ttttttgaga cagagtctca 11760
 ctctgtcgcc caggctggag tgcagtggca cgatctcggg tcaactgcaat ctccgcctc 11820
 ctgggttctt gctcagcct cccgagtagc tgggactaca ggcgcacgtc accacacccg 11880
 gctaattttt tgtattttta gtagtgacag ggattcaccg tottagccag ggtggtcttg 11940
 atctctgac ctcatgatct acccgctcgt gcttcccaaa gtgctgggat tacaggcatg 12000
 agccgctgtg cctggtctatt agagattttt tattataatt tatctccaag ataaaagcag 12060
 tgacattata ttgccacata attgaaaaat acaagagaaa taaaaatcat ccatgctttt 12120
 gttagcctat cactgtcatt gaaatattat gttacatggc agtttgcttg ctgggtgtct 12180
 tgttaggcaa cgctctgggt acattccttt agctattaat tgaggaatgt agaatgacag 12240
 aacagtgttt ctctcaatg atacttgaag gatatttatt aactttcata ttgaattaca 12300
 ttttattaaa tttataatga gttaatgctg ggaaataaaa cactgattta agtcattttg 12360
 gcttttagta ctaaagcatt tgacaataaa tgacttcttc agaatatggt atacctcttg 12420

aaagcaataa acgcatttta atgaattgta aggaaacaac atcattttat tttttatttt 12480
 tttttttgag acagactttc gcttttggtg cctaggtggt agtgcaatgg cggatctcog 12540
 gctcactgca acatccgcct ctgggttcaa gcgattctcc tgccctcagct tccctgagttg 12600
 ctgggattac aggcacgtgc caccacgcct ggccaatttt gtatttttag tagagacggg 12660
 gttttccac gtgggtcagg ctgggtctca actcctgacc tcagggtgac tgcccgcctc 12720
 agccctcgaa agtactggaa ttacaggcgt gagccaccgt gccctggcaa cattattatt 12780
 tttttttaat ctagaaaaat acacttctaa gaaaattgat taaaaccaac cttcttcatt 12840
 agccccaaag atcactctta tgttctcttt gttgcagctt cacttaaaaga gaggaaactt 12900
 cttggtgctc aaagcaaagg agatgggcct ccagtggtga gtgtgggggg taagctctct 12960
 ggggactcac tgggtacacc tgtccactta aggaatcac atttcacaga ggccctgcct 13020
 cttcatttca gtgggacagc tgccatogct cccatcattg ctgctgtcaa ggacgggaaa 13080
 agcatcactc atgaaggaa agagtgaga tgccctggtt tcttgatnca gcagttacag 13140
 gtagggtctg aaatgctggg cagagtctgt cttcttcagg cctacagac accacttttg 13200
 aaggacgtgg aacagtttgg acatcactca gctaagtgat aaaatggcct cttttatctg 13260
 tgtttgtccc gcatgtcaac acggctgcat tcgagcattt ttgtagattg tccatttagg 13320
 atctagtca cgtcctcctt aaagggtgca tgctttcctt ggtacttgag ctccagacag 13380
 tgtctaacaa cagaccccat atggatgggc ctgggggtta tggtcacag gaatgccaca 13440
 gtattctatg tcaagatatt tctctgact tctgaggaca ttaggaccag tggccacaga 13500
 ctgaagaaaa ccttaatgcc aagcctcctt tccctggccag tgtaggcctg aagtgcctca 13560
 acctgacagt tacctgttta ggtatocaca aagagaccag aagggtgttg atggtgatgt 13620
 gtaaaagtgg ttttgtgctt tgtttacctc tcagctcact ggaataggata tgtcatgtta 13680
 gcagttgcct tgaaggcagt tcagtttggg ggctgagctg tgacccccag tgggggggct 13740
 tatttggttt tgcagatttt ggctgaagag ctgtgtactc ctccagatcc tgggtgctgt 13800
 tttgtggtgg tagaatgtcc agatgaaagc ttcattcaac ccatctgtga gaatgccacc 13860
 tttcagaggt aatgaggggt ctctagggtg ggagaagtga gagctgaac ccagcccagc 13920
 atcgacatgg gcatcttggt gcaagagctg tgtttctggg aagaccacta totgggttta 13980
 cagttcagag gccggcactc ctgccttaag tcactgttgg tagttggtgg gctccggtgt 14040
 acacagccto aaagtgaat tagaaaagat tgaaaactag aaacaactga ggactagaaa 14100
 ttcaactaga actcttacag ctcttatacc agaagaaatt ctagaacttt tttgaattct 14160
 aactaatgcc ccagattatc atttggttta tttgaaactg aattaatttt cttccattac 14220

ctgcattgaa acaaatgagc tgggtcagag tgtgtgagac tgtcgtggte aagagtcggt 14280
 gttatgggat ggactcacag ctggggaatg tcttttgggc taactgccac tctgttgttg 14340
 tctctatcg aagttaacca gttttgcggt tcagctttca tccagatgg aatcatcttt 14400
 gaaccaccta tctgagtttg aatcttttcc cccactctta atggtttacc tgtatttttc 14460
 ctgttcttag tttgtatcta tctgtatttt ttcacttgtt tttttctact taccacaaca 14520
 aatccttttg ggctgctgta ccccttcoga gtcagagcgt taggagttgt ttcattgtct 14580
 gctttattct ctgtgggtga atttggaatg gctggtagcc ccggttttgt attttaatcc 14640
 agttttgggc agcaaaacct ctccaatgaa tcaggtgtca tttgagagcc atgtgtggat 14700
 gtgtgatgat gctgggtag ataaaaatag ctactgtgta tattttcttt taaagggaac 14760
 tggagggaaa cacatcagca tgttagtaag tggctgtgtg tccaggtggt gaaattttcag 14820
 atgattttca tttctcgtgc ctgtgtctca ggtcctctgg aaggcagaca ccagggtggc 14880
 attggagggt caggaggttt attcgaggaa atttgactgt gagagaggaa ggagagaggg 14940
 agcaggagga ggcaggggaga gcctgggtct ggctttgcag gttggaccg tatgagtga 15000
 gagggtagga aggaagtga gtgctgagaa aggatcagcc aggcctactg gaaagcccag 15060
 agcagagctt gccagataca ggaatccac gtcattgga aatggcccag caccggggtc 15120
 tgccgtgagc agcctgctgt gagagcatgg cctgggcgtg gaggtgtca gctcactgca 15180
 gtgctgcaga gggccgcacg ataccctccc ctggctgctg ggtccctgtc ttgggtgtgc 15240
 ctgagtctgc atcactttgt aaagcccac tcttctgccc aggtaccaag gaaagcgaga 15300
 tgcccccggt gccttggtgg ttcacatggc cccagcatct gtgctgtgtg acagcagga 15360
 ccagcagtg atggagaggt atggagccca gccagcggc acttggggta actcttctg 15420
 gcagtggttg attccccctt cctccccctg tgcctcttcc agcgtacct acccttctgc 15480
 acctgctaa actttctgtg ggaatcctgc cttcccagaa tttatagctt cccagatctg 15540
 tgcatactc gtgaagaaaa tgcaccgcta ggtggcgtag tgtccacag attccattta 15600
 ttttacccc tccacctct tcagggtgtc tgaacaaata ctgccgtttg gttgaggatt 15660
 ccataagtga attccaaaga agagattgca gctataaaat gatagcttcc atttactgaa 15720
 tgcccacttt gtgggagga gtgtgtgaaa tacccttcat ttcacttcat ttcctctagg 15780
 gtcgtgcaga gcagccctg gaggtagatg tttagtcaat ggaaggcata ttttctctg 15840
 gggcatcgct ggcaggggcc aggtggagga gtatgagttg agctcgggtg cgggggtgacc 15900
 ttgggtgct ttttgcccc tgccgtatc tccccacatg gcccgtttac ctgcccctca 15960
 ctccatggcc tgcctctctg ctgtctcttt cattcctcag gggttgggtc cctattttgt 16020

atgcacctgga catctctttt ttcttggttt tctctctact ctccaccga cacctgaaag 16080
 gcagctgagc tagggaacac cgggctttga gacagcagga gtgggacat gtttgccat 16140
 gtagtaacac tgcttggggc aagtcactga actgtttgaa cacctcatcc tcattaccac 16200
 tcttgagctc agcaccactc ctacggggga gctgcctctc aacagacgct gcaaatgccg 16260
 ggtctgttct ttcacagggt tgggcctgac acccagcact tggctcctgaa tgagaactgt 16320
 gcctcagttc acaacctctg cagccacaag attcaaaccc agctcaacct catccaccg 16380
 gacatcttcc cctgctccac cagtttccgc tgtaaggtag tgtctcagac yggcccttg 16440
 tcggccaccg tctcgtcccc tctctttctc tccatgaatg tgtttgtctc ctttcagaag 16500
 gagggcccca cctcagtggt gcccatgggt cagggtgaat gcctctccaa gtaccagctc 16560
 cgtccaccga gggagtggca gaggtctgtg ccatcttgaa ctaattggaat cgtctcagtc 16620
 gagttgggaa acatttctgt aaatagccac atagtaaatg ttccaggagg ctctccagac 16680
 catatggtct ctgttgtaac tattcaactc cgctttgagc aaaaagaaa caccgacaat 16740
 aagctaatag atgggcttgg ctgtgtgcca gctgaattt atttagaaaa gcagcctact 16800
 ccaggctggg ttgaggtggg cggattgggg ccagtagtct tcttttcca aaattgcctt 16860
 gcattgggaat agcagtata gagctcgtgt gtttcacagt atagaaaata ggaatgtgt 16920
 gatgaacaaa gtcaccata atctgtttgc ccagagataa tgattgataa cattttgtgt 16980
 ttcttgattt gtgtatgtg gtttatattg tcagttcttt cctgtatcac taaacagctc 17040
 taagtaacaa gatttttatt ggtattccaa atagggatgt ttactcattt gggatgttct 17100
 caattttttg ttgtttttaa tgaatgaaac aataaatgtc ttatatataa atctttgatg 17160
 ggaactctgt tccctcaag tcattcttaa atgtgggatt actggcccag agtgtgagac 17220
 ttgttaaggt acttgataaa tgtaagatgc catcttgaaa gcctcttcca gtacaatcca 17280
 accaggaaa tgacagcct tactgcccc catctttatt ttaattaatt aatttattta 17340
 tttttttt ttattttatt ttttgagacg gagtttact cttgttccc aggtggaggt 17400
 gcaatggcgt gatctcagct cactgcaacc tccgcctccc gggttcaagc atttctctg 17460
 ccccgacctc ccgaatagct ggaattacag gcgcctgcta ccacgcccgc ctaatttttt 17520
 gtaattttag cagagacggg ttccaccatg ttggcaggct ggtctcgaac tctgacctc 17580
 aggtgatcca cccacctcag cctcccaag tgttgggatt acaggcgtga gccgtgcccg 17640
 gcctgtttta atttttaagg atctgaacct tgattttaag ttctctgcc actccacagt 17700
 atttgtatta gaatagagca tgtgtgtgat tatgactgga tctgtgtgc tgttgaggtt 17760
 gggtagttgg ggccttttaa gagactatac tagcaagact cgggcccaca ggcaacatca 17820

cggggttgaa gaacotgggtg tccctttgtt ggcattctgcg caggctctta acacacagca 17880
 gcgatacaca gccctagccg acattcagat ttaccttgtg cttgtgaaaa atattgcaca 17940
 gggcctgccc tagacctagt gaattagaat cttgagagtt aggcttggga ctcacaagct 18000
 cccagatgat tttaatgtc agcgagggtg aagagccgcg tgtccaagga gttgccactc 18060
 cgtgtgatct ggggcttgct aggaaagtgg gatctcagcg ctcactgcag agctgocgaa 18120
 ctggcttctg cgttttgcca aggttccctg gtgtgaacat gagtctcaga gtcactcctc 18180
 tagggccctc gcttctcagc tcggaccatt gacccctcag aggacatttg gcaacatctg 18240
 gaaacgttct tgggtgtcac agcctaggag gtgggtagtgt gtgctgctag tgggttagag 18300
 tcaggggtag tgcaccagga cagcagcact ggccacagaa aaaaactgtc ttgccctgag 18360
 catcagtagt tccccgttga ctggccctga ggcagagcga tgcagcatcc aaaggcggt 18420
 ggagcagacc tgcgccagat cctagtcact taaccttcag tgttgatctg aaggaaactc 18480
 ctgcagattg tccccctgaa tttattcttg acatccccaa tggggctctg tgaggccata 18540
 taccctgtcc gtcacctgag atgcttctct ctctyctctg agggatgcca ttattacttg 18600
 caatcctgag gaattoatag ttgaggcgct gcagcttccc aacttcacag agagcgtgca 18660
 ggagtagcagg aggagtgcgc aggacggccc agccccagca ggtgagtggg agccacaga 18720
 gcagccttcc tttctcgggc tctgccctg ctgctgtttt cctagcatta agtggagtgc 18780
 tgggtggggc cattctaacc tggcttttca gtctaatac gggcttctct actcagctct 18840
 acattagaat tatagtcatt ggaggagggg gctttgggga gtttaagaat cccaattcct 18900
 ggetgggggc ggtggctcac acctgtaac ccagcacttc gggaggccga ggcaggtgga 18960
 tcgcgaggtc gggagatcga gaccatcctg gctaacatga tgaaaccccg tctctactaa 19020
 aaatacaaaa attagctggg cgtggcgggc ggcgcctgta gtcccagcta ctcgggagac 19080
 tgaggcagga gaatggcgag aaccggggag gcggagcttg cagtgcagca aggtcgtgcc 19140
 actgcgctcc agcctggagc acagagtgcg actccgtctc aaaaaaaaa atcccaattc 19200
 ctgtgcccca tcccaccaa tcagagcatt tggcgatggc acccaggcat tcttgcaag 19260
 gcacgcactg agtgaacgtc tttagtgaac acotgtggaa agagctctga gcagggaact 19320
 ggtcggcaga gatctagtcc tggctttgcg gatgcaaat catggaggat cttggccacg 19380
 tcaactcaat gaggctgagg gccgggcaca ggctttggaa ccactgggtc tccctggatt 19440
 tgaatcctga cctgcctct taccatcttc actggagacc tggcgctctg agcctgttcc 19500
 ccccttgga agcagagcat ttctacctg gtagggtctg gaggatgcga ccgaagtgc 19560
 tggctcttga gtgagagctg gatgcaaggc acacactgtt ctcttgaat aaatgacagt 19620

tcccagcata aagaaatgtc attttttaaa tgtaaaagaa ttacagcaat tcttttgaag 19680
 aaaggactgg agaatttatt tggtcttctt agccttttgg tgacagatag cctgtgggtc 19740
 ccacactggt gogaagtcct ttgtttcaga gcggttgcca ggggcctgcc agtccccctc 19800
 ctgggaagct ggatagaact atgttgctta cccatctgtc ttagtctgtg ttttgttatt 19860
 ataaaaaat atgtgagact gggtaattta tcaagaaaag aggtgtattt agttcacggt 19920
 tctgcgggct gagaattgaa ggtoacggcc ctagcttcca gtgaaggctt ccatgctgca 19980
 tcataacgtg gcagaggagc gcaagtagga agtgagcgtt tgtgaagacg ggataacctg 20040
 agctgcactc tggttttata acaacccccc totctggga acaaatccat tccttgaga 20100
 agtaatgcag tctcctgaga gccagtactt actactgcag ctccaagcca ctccaggagg 20160
 tccgtccctg tagcccaaac gccttcactt agggcccgcc tcccaaaacc gccatactag 20220
 ggagcagctt tccacatgag gtctggggac aaaccaatga cactcaaacc attgcacctt 20280
 ctcatggctg catgtggctt cacttttgac ccaaaggaat ggattgttcc acatggattt 20340
 tttcacagag aaaagaagtc agtaccaga aatcatcttc cttggaacag ggtctgccat 20400
 cccgatgaag attcgaaatg tcaagtccac acttgtcaac ataaggatg ctgctttccc 20460
 aggaagcacc ctccatcaa gggcaygttt actttttaa caaaagtccct gctgtactca 20520
 ccagtcgatt tgaatatgog tatcaagccc tgtcacttgt catgtcgact ggagtgacca 20580
 ggagaggagc gtggccttac tgcattttat agcctcagta gcaaacctta cctcggaat 20640
 caccaaaatt catcccatga tgtcttttaa taaacagctg attttactgt gggcagtaca 20700
 cctagctaag aaattagctc ctttaatttt tacattaatc ctatgaagtg gtgaataact 20760
 acccattttg ttgatgagtg acctgatatt cagagagggt acttgctatg gttcctacag 20820
 ctggtaagtg gggcatctga agtttgagcg gggacttggg gtcttgattg ctacatggta 20880
 ttgtccccc gccatttggt ggtagtatgt taaaaagctt taggggtttg cacatttggt 20940
 ttcagaacct ttattggatt ccccttgaca tgttttttag ttgattctct tgggtttgcc 21000
 tggggtoatc agcagagaga ttagtcaaat gcgttgtagc atgtacacgt tatctctaca 21060
 gatagtatgt gaagaaaata agattgtgaa ttaccaggtt tgtttttaat tttgctctgc 21120
 catcttacat gctagtgggt gatgataaac aaccaaatag tgcattaaat atatacagca 21180
 gtgacgagat gtgccctgac atcagaaaata tacaatctgg ggtgtgttcc tctgtggatg 21240
 aggacatgca ataaagcagc ttggagttag ccggcctctc ccgggggctg agatcctggg 21300
 ggaagaaggg ctttttgagt ttgacctgac accctgcgag cagcttttga accagctgaa 21360
 gctaattggga aggtgctatt gccaccttgc ctccgcctcc cgactccttt ttccccca 21420

aggtaatgtc ttgacccgg ggtctctctc tgcaaaatgg gtgcagccct ctcagtgttc 21480
 gtggctcctc ccagagaatg aaggaggcca gagcgggtca gcaactctctc tgccttgagg 21540
 cagagcttct gaaatggact gcacagcaga atagcccaag aagtttgca gaatccagac 21600
 ttccagagcc ctgcctaaaa ccaagtcaga aaccccgagt gacacctggg agtctgcgtt 21660
 aactggctcc ctgaatgaag cacctgcagc ccgcctgca ccaggtgtct ttgaggacat 21720
 gagctgagga aaccccgacc acttgcaaa ggggaaaagt ccgatggcag ctggacctag 21780
 aaagagtctc atatggccca gtgcctgtcc tggatttttc aacagaggct gtggccacag 21840
 tcaatctgca tggtcagatt cattgttagg actaaatgct ttaagcctcc tataaaattt 21900
 tttttttttt ttttgatgcc cagccttgt gtaagtctac ttgaaagggg ttcagggttc 21960
 catggatact tctttgctat aaagaggatg acacatgtaa aatcaccttt atggttaaat 22020
 taattggctt ttataattagc tctcaaaagc aaagcaggag agacagaaat ttctgcagtt 22080
 gottcttggg cctgtccaaa gcagacatca gcctctgaac catcagcagt ctctcagtg 22140
 gcagtgactc tcttctctt ctctctgca gccccgacac gtctctgcta ctggactgtg 22200
 gtgagggcac rtttgggcag ctgtgccgtc attacggaga ccaggtggac agggctcctgg 22260
 gcaccctggc tgctgtgtt gtgtcccacc tgcacgcaga tcaccacacg gtgagtgttg 22320
 ggctggacca caaagctgga gcctggagga ggcactgcca cgttgagttg gccctttggc 22380
 tgcgtctttt cctccgcttc caaacttgcc cagagctttt gttactcctc tctggctagg 22440
 aaatggtttt ttgcaaaact caacatagtc cttctgcgcc acaagaatgt cttctctcc 22500
 tgttcagttc ctttctgca gcagacagg tttgagtta ccagccttc cttgagtctt 22560
 gaatctcaca cggcctgtct agcggaaagt ttgaccgat gcaggaggtg tggctatgag 22620
 accctcaect tggctctctg gggtgccggg ccctgggccc ttgccctctt ccagcacagg 22680
 gtcgtgtcgc tttctgctg tgacatttca gggccatggc gcaggggggt cggcctgtgc 22740
 caccceact gcggtgtgt tagaggctgg tgggtgacgt cgggctggca actcctgcaa 22800
 gagagagggc tgcagaccct aacccgagg ggatggccct ggggcctggc tgacgcatgt 22860
 ctctgtttc cttgccagg cttgccaagt atcttgctgc agagagaacg cgccttggt 22920
 agtgtggcac ttgatggcg tcttgagttt cagcgtttta cacatcatcc gccatgctc 22980
 ttggcactcc agtttttatt gagatgttct gtctcgagt cggcacttgc attttttgtt 23040
 ccaggcatct ttgggaaagc cgttcaccc tttgtgtgtg gttgccccca accagctcaa 23100
 agcctggcto cagcagtacc acaaccagt ccaggaggtc ctgcaccaca tcaggtgagc 23160
 atccagggca gcctggccc stgggtgtt gcttgctgcc gtctccttca gaagctcaag 23220

gtggacaactg gggtagttac caatatcccc cagcagcett gcccttgaca tgggccaga 23280
 tggcagaagc aggggagaag tgcattggct gaaggacaga aaccattaga tagtteccat 23340
 gtaatgtcta ttttcttaga agcatttctt cccagtcctc atttgagttc tgagctgctt 23400
 tctaaacttc gagcagcttt tcttgatgag acagttccag agccaagcac ccaaatagtg 23460
 gctagcacag agaatgtcca tagcaggtgt gtggctagct ggcaggtggc accatcctca 23520
 ccccaagggg aaggagtccc ctctgctgga gccatccgtg gccctgtctg cctgagccgg 23580
 aggcagcatt cacctgctgg gtttctccca gtggcctaga ggctttgggt tggctcttta 23640
 tatttgactg ctgtttcttc atcatagtg ctatgattta actcatgttt tctcctaaga 23700
 atgatttttg ggttctccag ccaaagactt aaacttttgt tccagatgtc caagaaacct 23760
 ttattatcat tttaaatggt ttgtctttt acagtatgat tcttgccaaa tgccttcagg 23820
 aaggggctga gatctccagt cctgcagtgg aaagattgat cagtctgctg ttgcgaacrt 23880
 gtgattttga agaggttaagg ggcacagccg caggcatcat gggggcgagg tggggagcag 23940
 agctgcagag cctccagccc ccaccccttc agtttcagac ctgtctgggt cggcactgca 24000
 agcatgcgtt tggctgtgag ctgggtgcaca cctctggctg gaaagtggtc tattccgggg 24060
 acaccatgcc ctgcgaggt ctggtccgga tgggtgagta gaggaagaag caagccacc 24120
 tgaggttgct ctgggggttg ttagctgga ggtgaatgca ggtgggcttg cagggaaacg 24180
 tcagcagagg caggagactc aggtccccc cctcagagtc tctggttgct atcctagtag 24240
 gcagaccagg gccaggggga gctgagtgt gagaccagga aacgacagct gactgaggcc 24300
 tgtgtgccgc tctcgagag aactctgcc tgatccttgt gctgcttctc cagggaaaag 24360
 tggccacctc ctgatacatg aagccacct ggaagatggt ttggaagagg aagcagtgga 24420
 aaagacacac aggtagcaaa gggcgggtcag tccttgctgc ccacatctc tccctcccc 24480
 actacgtgac actgagcagc cgttgtttgt ctccactgat gtggggctgc cctgcttct 24540
 atcaagggtg atgggggctt ccttgacctg tggcagtgt caccaggtct tggcctttat 24600
 ttttcagaa ttttctaagc aagattctag agtgaggcac agttttttga aagcatctag 24660
 aaatcggctg aataaactat aagccatgtc agggaattgc caggggaagg cgggggctgg 24720
 gggactgaat ttttgctgc taatttcaac gaaagagtgc attaccccag gtgggcccctg 24780
 tggtttctct tgggtgccct catggacaga ttggcgagcc agcacagagg gtggggtcca 24840
 tccaggggtg tgtgcgaagg ctctggccct caggggagat tgtgctggct acggaggtgc 24900
 ccgttaagaa aaccaccag ctccccggg tgccctggca gttgatggcc agggctctgtg 24960
 ccactgtctg ctttgcatgc ttgcagtga gttcagcttc agtctgtctc gtccttcacc 25020

tgcagcaca cgtcccaagc catcagcgtg gggatgcgga tgaacgcgga gttcattatg 25080
 ctgaaccact tcagccagcg ctatgccaaag gtccccctct tcagcccaa cttcagcgag 25140
 aaagtgggag ttgcctttga ccacatgaag gtctgtatgt cacacggaca gcacagggag 25200
 gggacggggc agggagacag gactctacac actgagtagg acggtcagct ggagtttctg 25260
 ttctattttg gggccaccgt gggaaaaggt tatctacca tcaactaacc ggtcgaacca 25320
 ccctgggttt gctggtgaga cccacctcct gcaggggcca actagtcttc agtctcagtt 25380
 cactggaat ttctgagaat ctttttaggc ctgactgct cacacagtca tggcatttga 25440
 gcctcagcac agacctgtga gacaggtgtg tgcctcttgt gagtgggaaa gccaggcctg 25500
 acccttgccc ttccggaatg aaggggcaga gccggagcca ggctcgttt ttcaggagct 25560
 tgatttttag agcatctgga ctgctctccc ttccctctcc ggaggccctt agccaggcct 25620
 ggggagcctc tgcctcttta gagggttccc tccatgccat tctttttccc atttcagctg 25680
 tggcctgttg gcttgtgcc aaggaagggc gttggcgctg ctgtgtgagc acatgactgc 25740
 atcccttcca gctcctgtcc cccaccctg cccctctgag acatgtcctt gtctcttatt 25800
 gtgtcttcta ggtctgcttt ggagactttc caacaatgcc caagctgatt cccccactga 25860
 aagccctgtt tgctggcgac atcagaggaga tggaggagcg cagggagaag cgggagctgc 25920
 ggcaggtgcg ggcggccctc ctgtccaggg agctggcagg cggcctggag gatggggagc 25980
 ctcagcagaa gcgggccacc acagaggagc cacaggccaa gaaggtcaga gcccagtga 26040
 gatctgggag accctgaact cagaagcgtg tgtgtcttct gccccacgca cgcaccctga 26100
 tctgccctcc ttgctggtag aagctgaaga gcacggctcc ccaggaggca gctcaggata 26160
 ggtggtatgg agctgtgccg aggcttgggc tcccacataa gcaactagtct atagatgcct 26220
 cttaggactg gtgcctggca cagccgcggg ccaggaggct gccacacgga agcaagcaga 26280
 tgaactaatt tcatttcaag gcagttttta aagaagctctt ggaacacagc gccggcacct 26340
 ttctctata ccagcaaaagt gattccctgc acaccagaga caagcagagt aacaggatca 26400
 gtgggtctaa gtgtccgaga cttaacgaaa atagtatttc agctgcaata aagattgagt 26460
 ttgcaattgt gatttctttt gtttctctct gctgtgctca cagagcaggg tctgctgtgc 26520
 accaccttgg agaaggctct ctgtgctgta gtgtggcagc tgccctggta cccgggtgct 26580
 tggaagaagt cagctccctg cgtagtgcgc acctctggaa cctgtctcca gagagccacc 26640
 cttattcgcc aagtcttttt gaca 26664

<210> 29
 <211> 31
 <212> DNA

<213> Homo sapiens

<400> 29
caggaattca gcacatactc attgttcagn n 31

<210> 30

<211> 21

<212> DNA

<213> Homo sapiens

<400> 30
caggaattca gcacatactc a 21

<210> 31

<211> 22

<212> DNA

<213> Homo sapiens

<400> 31
ttcagcacat actcattgtt ca 22

<210> 32

<211> 19

<212> DNA

<213> Homo sapiens

<400> 32
tgaacgcctt ctccacagt 19

<210> 33

<211> 17

<212> DNA

<213> Homo sapiens

<400> 33
gtaccgcgtg ccaccac 17

<210> 34

<211> 33

<212> DNA

<213> Homo sapiens

<400> 34
gctaggatcc gccaccatgt gggcgctttg ctc 33

<210> 35

<211> 29

<212> DNA

<213> Homo sapiens

<400> 35
gtactctgag tcaactgggct ctgaccttc 29

<210> 36
 <211> 17
 <212> DNA
 <213> Homo sapiens

<400> 36
 gtaaaacgac ggccagt 17

<210> 37
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 37
 ggaaacagct atgaccatg 19

<210> 38
 <211> 17
 <212> DNA
 <213> Homo sapiens

<400> 38
 tgccgacgcg agagaag 17

<210> 39
 <211> 17
 <212> DNA
 <213> Homo sapiens

<400> 39
 cgcttctctc gcgtgcg 17

<210> 40
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 40
 tctaagtgtg ggggctta 18

<210> 41
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 41
 taagcccccacattaga 18

<210> 42
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 42

tgaaaatgag ccacacct

18

<210> 43
<211> 18
<212> DNA
<213> Homo sapiens

<400> 43
aggtgtggct cattttca

18

<210> 44
<211> 18
<212> DNA
<213> Homo sapiens

<400> 44
cattcaaccc atctgtga

18

<210> 45
<211> 18
<212> DNA
<213> Homo sapiens

<400> 45
tcacagatgg gttgaatg

18

<210> 46
<211> 18
<212> DNA
<213> Homo sapiens

<400> 46
tgaatgcctc ctcaagta

18

<210> 47
<211> 18
<212> DNA
<213> Homo sapiens

<400> 47
tacttgagga ggcatcca

18

<210> 48
<211> 18
<212> DNA
<213> Homo sapiens

<400> 48
gctactggac tgtggtga

18

<210> 49
<211> 18
<212> DNA

<213> Homo sapiens

<400> 49

tcaccacagt ccagtagc

18

<210> 50

<211> 19

<212> DNA

<213> Homo sapiens

<400> 50

tggaagagtt tcagacctg

19

<210> 51

<211> 19

<212> DNA

<213> Homo sapiens

<400> 51

caggtctgaa actcttcca

19

<210> 52

<211> 17

<212> DNA

<213> Homo sapiens

<400> 52

cgcagggacg caccata

17

<210> 53

<211> 18

<212> DNA

<213> Homo sapiens

<400> 53

ggttgaactc ggagaaga

18

<210> 54

<211> 19

<212> DNA

<213> Homo sapiens

<400> 54

caactggaaa aatacctcg

19

<210> 55

<211> 17

<212> DNA

<213> Homo sapiens

<400> 55

gcagagtcca gaaagcg

17

<210> 56
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 56
 agaggaaact tcttggtgc 19

<210> 57
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 57
 accaaggaaa ggcagatg 18

<210> 58
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 58
 gtcaacataa gccccgac 18

<210> 59
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 59
 ggetgctgtg tttgtgtc 18

<210> 60
 <211> 17
 <212> DNA
 <213> Homo sapiens

<400> 60
 gaaggcattt ggcagga 17

<210> 61
 <211> 18
 <212> DNA
 <213> Homo sapiens

<400> 61
 tatgattcct gccaaatg 18

<210> 62
 <211> 17
 <212> DNA
 <213> Homo sapiens

<400> 62

tccagccaga ggtgtgc

17

<210> 63
<211> 17
<212> DNA
<213> Homo sapiens

<400> 63
tgcgaggctc tgggtccg

17

<210> 64
<211> 18
<212> DNA
<213> Homo sapiens

<400> 64
gggcattggt ggaaagtc

18

<210> 65
<211> 17
<212> DNA
<213> Homo sapiens

<400> 65
tgtttgetgg cgacatc

17

<210> 66
<211> 31
<212> DNA
<213> Homo sapiens

<400> 66
caggaattca gcacatactc attgttcagn n

31

<210> 67
<211> 21
<212> DNA
<213> Homo sapiens

<400> 67
caggaattca gcacatactc a

21

<210> 68
<211> 22
<212> DNA
<213> Homo sapiens

<400> 68
ttcagcacat actcattggt ca

22

<210> 69
<211> 19
<212> DNA

<213> Homo sapiens

<400> 69

cagaacacat ttgggaagc

19

<210> 70

<211> 18

<212> DNA

<213> Homo sapiens

<400> 70

gatgttgtcc aagcgagc

18

<210> 71

<211> 17

<212> DNA

<213> Homo sapiens

<400> 71

tgacacacag cacctga

17

<210> 72

<211> 17

<212> DNA

<213> Homo sapiens

<400> 72

gaagatgtca ggggtgga

17

<210> 73

<211> 18

<212> DNA

<213> Homo sapiens

<400> 73

caggcatacc actacaga

18

<210> 74

<211> 20

<212> DNA

<213> Homo sapiens

<400> 74

tatcaacttc taggcaagtg

20

<210> 75

<211> 19

<212> DNA

<213> Homo sapiens

<400> 75

gcaccatgtc gcagggttc

19

<210> 76
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 76
 gaaccctgcg acatgggtgc 19

<210> 77
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 77
 tcgcaggggtt cggctcgtc 19

<210> 78
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 78
 aaccctgcga catgggtgcg 19

<210> 79
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 79
 aaagaccac tgcgacacc 19

<210> 80
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 80
 gcaggtgtcg cagtgggtc 19

<210> 81
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 81
 ccgaacaccg tgtacctgca 20

<210> 82
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 82

cagggtacacg gtgttcggg

19

<210> 83
<211> 21
<212> DNA
<213> Homo sapiens

<400> 83
gtcttctcgg aatacaacag g

21

<210> 84
<211> 20
<212> DNA
<213> Homo sapiens

<400> 84
ctgttggtatt ccgagaagac

20

<210> 85
<211> 20
<212> DNA
<213> Homo sapiens

<400> 85
aaggcggtcca acgacttatg

20

<210> 86
<211> 20
<212> DNA
<213> Homo sapiens

<400> 86
agtcgttgga cgccttctcc

20

<210> 87
<211> 20
<212> DNA
<213> Homo sapiens

<400> 87
tccgagtcag aaagatgttg

20

<210> 88
<211> 17
<212> DNA
<213> Homo sapiens

<400> 88
gccttggtcag cctgggtg

17

<210> 89
<211> 17
<212> DNA

<213> Homo sapiens

<400> 89

aggaagtgag cagagcgc

17

<210> 90

<211> 33

<212> DNA

<213> Homo sapiens

<400> 90

gctaaagcct gccaccatgt gggcgctccg ctc

33

<210> 91

<211> 27

<212> DNA

<213> Homo sapiens

<400> 91

gctactcgag tcacactcgc gctccta

27

<210> 92

<211> 16

<212> DNA

<213> Homo sapiens

<400> 92

gccttctccg cagtta

16

<210> 93

<211> 21

<212> DNA

<213> Homo sapiens

<400> 93

ccgcctgaga cgctctagta t

21

<210> 94

<211> 19

<212> DNA

<213> Homo sapiens

<400> 94

gctccgaaag tgctgacag

19

<210> 95

<211> 39

<212> DNA

<213> Homo sapiens

<400> 95

gttttcccag tcacgacgtt tctattggat gagcagcct

39

<210> 96
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 96
 aggaaacagc tatgaccatg cctgcgatat ggtgcgtc

38

<210> 97
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 97
 gttttcccag tcacgacgct cagttttggt ggagacg

37

<210> 98
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 98
 aggaaacagc tatgaccatg tgccccgatg ctcagag

37

<210> 99
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 99
 aatgggtgtca gagagtttac ag

22

<210> 100
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 100
 gctatttggg aggctgagg

19

<210> 101
 <211> 40
 <212> DNA
 <213> Homo sapiens

<400> 101
 gtttttcccag tcacgacgaa tgggtgtcaga gagtttacag

40

<210> 102
 <211> 41
 <212> DNA
 <213> Homo sapiens

<400> 102

aggaaacagc tatgaccatg aacaaggacc acttttgcta t

41

<210> 103

<211> 40

<212> DNA

<213> Homo sapiens

<400> 103

gttttcccag tcacgacgtt tatagcaaaa gtggtccttg

40

<210> 104

<211> 38

<212> DNA

<213> Homo sapiens

<400> 104

aggaaacagc tatgaccatg agacttccca ccagcctc

38

<210> 105

<211> 19

<212> DNA

<213> Homo sapiens

<400> 105

ccttgctgct tcaccctag

19

<210> 106

<211> 21

<212> DNA

<213> Homo sapiens

<400> 106

tgcttttatat gtgctgctac g

21

<210> 107

<211> 39

<212> DNA

<213> Homo sapiens

<400> 107

gttttcccag tcacgacgca tcttccctgg ttgtacttc

39

<210> 108

<211> 39

<212> DNA

<213> Homo sapiens

<400> 108

aggaaacagc tatgaccatc tggagggcag aagactgat

39

<210> 109

<211> 23

<212> DNA

<213> Homo sapiens

<400> 109

ctacatttgt tcaaccataa ctg

23

<210> 110

<211> 23

<212> DNA

<213> Homo sapiens

<400> 110

gattttgagg ttgatgttg atg

23

<210> 111

<211> 39

<212> DNA

<213> Homo sapiens

<400> 111

gttttccag tcacgacga ttgttcaac cataactgc

39

<210> 112

<211> 39

<212> DNA

<213> Homo sapiens

<400> 112

aggaaacagc tatgaccata ttgagaggt cagggcata

39

<210> 113

<211> 21

<212> DNA

<213> Homo sapiens

<400> 113

tcgtgtcaga ttcccacat a

21

<210> 114

<211> 21

<212> DNA

<213> Homo sapiens

<400> 114

aggcataagt cagacatccg t

21

<210> 115

<211> 40

<212> DNA

<213> Homo sapiens

<400> 115

gttttccag tcacgacggt tactcttccc acacatcttc

40

<210> 116
 <211> 41
 <212> DNA
 <213> Homo sapiens

 <400> 116
 aggaaacagc tatgaccatc acagcaagtg ttcagtttct a 41

 <210> 117
 <211> 21
 <212> DNA
 <213> Homo sapiens

 <400> 117
 cattcccatg tatgaacgtc t 21

 <210> 118
 <211> 22
 <212> DNA
 <213> Homo sapiens

 <400> 118
 atagtaagcc caggaagaag ga 22

 <210> 119
 <211> 39
 <212> DNA
 <213> Homo sapiens

 <400> 119
 gttttcccag tcacgacgca ttcccatgta tgaacgtct 39

 <210> 120
 <211> 41
 <212> DNA
 <213> Homo sapiens

 <400> 120
 aggaaacagc tatgaccatc tacaagcatt acaaggcaga g 41

 <210> 121
 <211> 22
 <212> DNA
 <213> Homo sapiens

 <400> 121
 agtgtcttca gcctttgtat tg 22

 <210> 122
 <211> 23
 <212> DNA
 <213> Homo sapiens

 <400> 122

atctgctatc tctttctgtc tca

23

<210> 123
<211> 40
<212> DNA
<213> Homo sapiens

<400> 123
gttttccag tcacgacgat cgggtcataa tcagtctgtg

40

<210> 124
<211> 42
<212> DNA
<213> Homo sapiens

<400> 124
aggaaacagc tatgaccata tcttctcttg tctcaggtaa ca

42

<210> 125
<211> 22
<212> DNA
<213> Homo sapiens

<400> 125
cttctgaaag caataaacgc at

22

<210> 126
<211> 20
<212> DNA
<213> Homo sapiens

<400> 126
gatgtccaaa ctgttccacg

20

<210> 127
<211> 40
<212> DNA
<213> Homo sapiens

<400> 127
gttttccag tcacgacgta aaaccaacct tcttcattag

40

<210> 128
<211> 38
<212> DNA
<213> Homo sapiens

<400> 128
agggaaacagc tatgaccata gcaatgatgg gagcgatg

38

<210> 129
<211> 37
<212> DNA

<213> Homo sapiens

<400> 129

gttttcccag tcacgacggg cttctgggga ctcaactg

37

<210> 130

<211> 41

<212> DNA

<213> Homo sapiens

<400> 130

aggaaacagc tatgaccatc cttcaaaagt ggtgtctgta g

41

<210> 131

<211> 22

<212> DNA

<213> Homo sapiens

<400> 131

gtatccacaa agagaccaga ag

22

<210> 132

<211> 23

<212> DNA

<213> Homo sapiens

<400> 132

caccaactac caacagtgac tta

23

<210> 133

<211> 41

<212> DNA

<213> Homo sapiens

<400> 133

gttttcccag tcacgacggc tcaactggata ggatatgtca t

41

<210> 134

<211> 39

<212> DNA

<213> Homo sapiens

<400> 134

aggaaacagc tatgaccatc cagaaacaca gctcttgcc

39

<210> 135

<211> 20

<212> DNA

<213> Homo sapiens

<400> 135

gcttgccaga tacaggaatc

20

<210> 136
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 136
 acagaaagtt taggcaggtg 20

<210> 137
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 137
 gttttccag tcacgacgac gatacccctc cctggct 37

<210> 138
 <211> 39
 <212> DNA
 <213> Homo sapiens

<400> 138
 aggaaacagc tatgaccata cagaaagttt aggcaggtg 39

<210> 139
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 139
 cctctcactc ttcccagcac 20

<210> 140
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 140
 ggagtaggct gcttttctaa at 22

<210> 141
 <211> 40
 <212> DNA
 <213> Homo sapiens

<400> 141
 gttttccag tcacgacgga acacctcadc ctcattacca 40

<210> 142
 <211> 41
 <212> DNA
 <213> Homo sapiens

<400> 142

aggaacacagc tatgaccata agagacaaaa cacattcatg g

41

<210> 143

<211> 39

<212> DNA

<213> Homo sapiens

<400> 143

gttttccag tcacgacggt ttccgctgta aggtagtgt

39

<210> 144

<211> 42

<212> DNA

<213> Homo sapiens

<400> 144

aggaacacagc tatgaccatc tggacattt actatgtggc ta

42

<210> 145

<211> 20

<212> DNA

<213> Homo sapiens

<400> 145

tgctagtggg tagaggctcag

20

<210> 146

<211> 21

<212> DNA

<213> Homo sapiens

<400> 146

actgaaagcc aggttagaat g

21

<210> 147

<211> 37

<212> DNA

<213> Homo sapiens

<400> 147

gttttccag tcacgacgac cctgtccgtc acctgag

37

<210> 148

<211> 38

<212> DNA

<213> Homo sapiens

<400> 148

aggaacacagc tatgaccatc ccaccagcac tccactta

38

<210> 149

<211> 21

<212> DNA

<213> Homo sapiens

<400> 149

tgtgaagacg ggataacctg a

21

<210> 150

<211> 19

<212> DNA

<213> Homo sapiens

<400> 150

gacagggett gataccgca

19

<210> 151

<211> 38

<212> DNA

<213> Homo sapiens

<400> 151

gttttcccag tcacgacgat gctggctcac ttttgacc

38

<210> 152

<211> 39

<212> DNA

<213> Homo sapiens

<400> 152

aggaaacagc tatgaccatg actggtgagt acagcagga

39

<210> 153

<211> 21

<212> DNA

<213> Homo sapiens

<400> 153

ccagcctttg tgtaagtcta c

21

<210> 154

<211> 19

<212> DNA

<213> Homo sapiens

<400> 154

tctgggcaag tttggaagc

19

<210> 155

<211> 39

<212> DNA

<213> Homo sapiens

<400> 155

gttttcccag tcacgacgtc caaagcagac atcagcctc

39

<210> 156
 <211> 38
 <212> DNA
 <213> Homo sapiens

<400> 156
 aggaaacagc tatgaccatg gaggaaga cgcagcca

38

<210> 157
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 157
 cgctttctgc ctgtgacat

19

<210> 158
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 158
 ttctgtcctt cagccaatgc

20

<210> 159
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 159
 gttttccag tcacgacgtt agaggctggt gggtgac

37

<210> 160
 <211> 42
 <212> DNA
 <213> Homo sapiens

<400> 160
 aggaaacagc tatgaccatc atctcaataa aaactggagt gc

42

<210> 161
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 161
 cacttgatgg gcgttctgag

20

<210> 162
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 162

ttctgtcctt cagccaatgc

20

<210> 163

<211> 38

<212> DNA

<213> Homo sapiens

<400> 163

gttttccag tcacgacgtt ccagcggttt acacatca

38

<210> 164

<211> 38

<212> DNA

<213> Homo sapiens

<400> 164

aggaaacagc tatgaccatt accccagtg ccaccttg

38

<210> 165

<211> 20

<212> DNA

<213> Homo sapiens

<400> 165

gggttctcca gccaaagact

20

<210> 166

<211> 19

<212> DNA

<213> Homo sapiens

<400> 166

ctgagttctcc tgcctctgc

19

<210> 167

<211> 38

<212> DNA

<213> Homo sapiens

<400> 167

gttttccag tcacgacggg gttctcagc caaagact

38

<210> 168

<211> 38

<212> DNA

<213> Homo sapiens

<400> 168

aggaaacagc tatgaccatg tggggctgga aggctctg

38

<210> 169

<211> 37

<212> DNA

<213> Homo sapiens

<400> 169
gttttcccag tcacgacgaa gaggtaaggg gcacagc 37

<210> 170

<211> 38

<212> DNA

<213> Homo sapiens

<400> 170
aggaaacagc tatgaccatc tgagtctcct gcctctgc 38

<210> 171

<211> 20

<212> DNA

<213> Homo sapiens

<400> 171
gctgagtgtt gagaccagga 20

<210> 172

<211> 19

<212> DNA

<213> Homo sapiens

<400> 172
agacaaacga cggctgctc 19

<210> 173

<211> 38

<212> DNA

<213> Homo sapiens

<400> 173
gttttcccag tcacgacgtt gagaccagga aacagcac 38

<210> 174

<211> 38

<212> DNA

<213> Homo sapiens

<400> 174
aggaaacagc tatgaccatg agaggatgtg ggcgacaa 38

<210> 175

<211> 19

<212> DNA

<213> Homo sapiens

<400> 175
gggagatggg gctggctac 19

<210> 176
 <211> 21
 <212> DNA
 <213> Homo sapiens

<400> 176
 cctgggttagt gatgggtaga t 21

<210> 177
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 177
 gttttccag tcacgacgca gggctotgtgc cactgtc 37

<210> 178
 <211> 40
 <212> DNA
 <213> Homo sapiens

<400> 178
 aggaaaacagc tatgaccatc tcagtgtgta gagtcctctc 40

<210> 179
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 179
 ttgattttga gagcatctgg ac 22

<210> 180
 <211> 21
 <212> DNA
 <213> Homo sapiens

<400> 180
 ctcggacact tagaccact g 21

<210> 181
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 181
 gttttccag tcacgacgtg catcccttcc agctcct 37

<210> 182
 <211> 41
 <212> DNA
 <213> Homo sapiens

<400> 182

aggaaacagc tatgaccatg acacacagcc ttotgagttc a

41

<210> 183

<211> 37

<212> DNA

<213> Homo sapiens

<400> 183

gttttcccag tcacgacgcc acacagagga gccacag

37

<210> 184

<211> 42

<212> DNA

<213> Homo sapiens

<400> 184

aggaacacgc tatgaccata ccagtcctaa gaggcattta ta

42

<210> 185

<211> 19

<212> DNA

<213> Homo sapiens

<400> 185

ccacacagag gagccacag

19

<210> 186

<211> 20

<212> DNA

<213> Homo sapiens

<400> 186

ccagagggtgc tcactacgac

20

<210> 187

<211> 39

<212> DNA

<213> Homo sapiens

<400> 187

gttttcccag tcacgacgag gtcagagccc agtgaagat

39

<210> 188

<211> 39

<212> DNA

<213> Homo sapiens

<400> 188

aggaacacgc tatgaccatc atctgcttgc ttccgtgtg

39

<210> 189

<211> 39

<212> DNA

<213> Homo sapiens

<400> 189
gttttcccag tcacgacgtc aggatagggtg gtatggagc 39

<210> 190

<211> 40

<212> DNA

<213> Homo sapiens

<400> 190
aggaacacgc tatgaccatc ggacacttag acccactgat 40

<210> 191

<211> 23

<212> DNA

<213> Homo sapiens

<400> 191
agactccgag tygaatgaaa atg 23

<210> 192

<211> 23

<212> DNA

<213> Homo sapiens

<400> 192
ggtgagggca crtttgggca gct 23

<210> 193

<211> 23

<212> DNA

<213> Homo sapiens

<400> 193
gcaccctggc trectgtgttt gtg 23

<210> 194

<211> 23

<212> DNA

<213> Homo sapiens

<400> 194
gtgtcccacc tgcacgcaga tca 23

<210> 195

<211> 24

<212> DNA

<213> Homo sapiens

<400> 195
gtgtcccacc tggcagcgag atca 24

<210> 196
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 196
 aagccgcttc aycctttgct ggt

23

<210> 197
 <211> 24
 <212> DNA
 <213> Homo sapiens

<400> 197
 gctgttgcca acrtgtgatt tgga

24

<210> 198
 <211> 22
 <212> DNA
 <213> Homo sapiens

<400> 198
 gaggccttggg stccacata ag

22

<210> 199
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 199
 cctggcacag cygcgggcca gga

23

<210> 200
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 200
 aatccagcaa artgattccc tgc

23

<210> 201
 <211> 21
 <212> DNA
 <213> Homo sapiens

<400> 201
 taaatgtttt ytcattctta g

21

<210> 202
 <211> 23
 <212> DNA
 <213> Homo sapiens

<400> 202

ttgctgttgt gyggttttct tgt

23

<210> 203

<211> 24

<212> DNA

<213> Homo sapiens

<400> 203

ggttttcttg attcagcagt taca

24

<210> 204

<211> 27

<212> DNA

<213> Homo sapiens

<400> 204

ggttttcttg atgattcagc agttaca

27

<210> 205

<211> 23

<212> DNA

<213> Homo sapiens

<400> 205

gtgtctcaga cyggccctt gtc

23

<210> 206

<211> 23

<212> DNA

<213> Homo sapiens

<400> 206

tgccatcttg awctaataatga atc

23

<210> 207

<211> 23

<212> DNA

<213> Homo sapiens

<400> 207

cttctctctc tyoctgcagg gat

23

<210> 208

<211> 23

<212> DNA

<213> Homo sapiens

<400> 208

catcaagggc aygtttactt ttt

23

<210> 209

<211> 23

<212> DNA

<213> Homo sapiens

<400> 209
cagccttgcc cscctgggctg ttg

23

<210> 210

<211> 350

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (51)..(293)

<400> 210

cgcgggcgta ggtgaccggc ggcttttctca gttttggtgg agacggggcgc atg tgg 56
Met Trp
1

gcg ctt tgc tcg ctg ctg cgg tcc gcg gcc gga cgc acc atg tcg cag 104
Ala Leu Cys Ser Leu Leu Arg Ser Ala Ala Gly Arg Thr Met Ser Gln
5 10 15

gga cgc acc ata tcg cag gca ccc gcc cgc cgc gag cgg ccg cgc aag 152
Gly Arg Thr Ile Ser Gln Ala Pro Ala Arg Arg Glu Arg Pro Arg Lys
20 25 30

gac cgg ctg cgg cac ctg cgc acg cga gag aag cgc gga ccg tcg ggg 200
Asp Pro Leu Arg His Leu Arg Thr Arg Glu Lys Arg Gly Pro Ser Gly
35 40 45 50

tgc tcc ggc ggc cca aac acc gtg tac ctg cag gtg gtg gca gcg ggt 248
Cys Ser Gly Gly Pro Asn Thr Val Tyr Leu Gln Val Val Ala Ala Gly
55 60 65

agc cgg gac tcg ggc gcc gcg ctc tac gtc ttc tcc gag ttc aac 293
Ser Arg Asp Ser Gly Ala Ala Leu Tyr Val Phe Ser Glu Phe Asn
70 75 80

cggtcagtc acgagccacg ccccgccccg ctgggcccctc agtgcgggcgc agcctct 350

<210> 211

<211> 81

<212> PRT

<213> Homo sapiens

<400> 211

Met Trp Ala Leu Cys Ser Leu Leu Arg Ser Ala Ala Gly Arg Thr Met
1 5 10 15

Ser Gln Gly Arg Thr Ile Ser Gln Ala Pro Ala Arg Arg Glu Arg Pro
20 25 30

Arg Lys Asp Pro Leu Arg His Leu Arg Thr Arg Glu Lys Arg Gly Pro
35 40 45

Ser Gly Cys Ser Gly Gly Pro Asn Thr Val Tyr Leu Gln Val Val Ala
50 55 60

Ala Gly Ser Arg Asp Ser Gly Ala Ala Leu Tyr Val Phe Ser Glu Phe
 65 70 75 80

Asn

<210> 212
 <211> 326
 <212> DNA
 <213> Mus musculus

<220>
 <221> CDS
 <222> (51)..(269)

<400> 212
 tggcggcggtg aggggtcttg ctgccttgct agcctgggtgt ggtcgggtgc atg tgg 56
 Met Trp
 1
 ggc ctc cgc tca ctg ttg cgt ccc ctt ggc ctg cgc acc atg tgc cag 104
 Ala Leu Arg Ser Leu Leu Arg Pro Leu Gly Leu Arg Thr Met Ser Gln
 5 10 15
 ggt tgc gct cgt cgg cgg cgg cca ccc aag gac cca ctg cga cac ctg 152
 Gly Ser Ala Arg Arg Pro Arg Pro Pro Lys Asp Pro Leu Arg His Leu
 20 25 30
 cgt acg cgg gag aag cgc ggc cgg ggt ccc ggg ggc cgg aac acc gtg 200
 Arg Thr Arg Glu Lys Arg Gly Pro Gly Pro Gly Gly Pro Asn Thr Val
 35 40 45 50
 tac ctg cag gtg gtg gcg gcg ggc ggc cgg gac gcg ggg gct gct ctc 248
 Tyr Leu Gln Val Val Ala Ala Gly Gly Arg Asp Ala Gly Ala Ala Leu
 55 60 65
 tat gtc ttc tgc gaa tac aac aggtcagagt gggccgacag ccctggggga 299
 Tyr Val Phe Ser Glu Tyr Asn
 70
 ttggccccag cgccacgtgc tcgggag 326

<210> 213
 <211> 73
 <212> PRT
 <213> Mus musculus

<400> 213
 Met Trp Ala Leu Arg Ser Leu Leu Arg Pro Leu Gly Leu Arg Thr Met
 1 5 10 15
 Ser Gln Gly Ser Ala Arg Arg Pro Arg Pro Pro Lys Asp Pro Leu Arg
 20 25 30
 His Leu Arg Thr Arg Glu Lys Arg Gly Pro Gly Pro Gly Gly Pro Asn
 35 40 45

Thr	Val	Tyr	Leu	Gln	Val	Val	Ala	Ala	Gly	Gly	Arg	Asp	Ala	Gly	Ala
	50						55				60				
Ala	Leu	Tyr	Val	Phe	Ser	Glu	Tyr	Asn							
65					70										

62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80